

SEARCH REQUEST FORM

Requestor's Name: _____ Serial Number: _____
Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 05-13-03
Searcher: Beverly C4994
Terminal time: 20
Elapsed time: _____
CPU time: _____
Total time: 23
Number of Searches: _____
Number of Databases: 1

Search Site

_____ STIC
_____ CM-1
_____ Pre-S

Type of Search

_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure
_____ Bibliographic

Vendors

_____ IG Suite
_____ STN
_____ Dialog
_____ APS
_____ Geninfo
_____ SDC
_____ DARC/Questel
✓ Other CGN

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 11:30:07 ; Search time 80 seconds
(without alignments)
1422.450 million cell updates/sec

Title: US-09-830-762-2
Perfect score: 4447
Sequence: 1 MDFSVMAGAAVNEKSGRT.....PMRPAQHPWRGASGSPVD 854

Scoring table: BLOSUM62
Gap 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4447	100.0	854	21	AAV71082
2	4208.5	94.6	819	23	ABR05392
3	3927.5	88.3	780	23	ABR05391
4	3388	76.2	653	22	ABR50253
5	2453	55.2	826	21	AAV71083
6	1888	42.5	360	22	AAV79380
7	1765	39.7	335	22	AAV78396
8	1104	24.8	249	23	ABR41342
9	420	9.4	556	22	ABR93319
10	363	8.2	444	22	AAE01176

11	363	8.2	444	23	ABG64060	Human albumin fusi
12	291.5	6.6	599	23	ABR84807	Thermotoga maritima
13	256	5.8	419	22	ABR94051	Human protein sequ
14	235	5.3	176	23	ABR93353	Listeria monocytog
15	219	4.9	246	22	AAU58306	Protonibacterium
16	211.5	4.8	1025	22	AAU58389	Human protein sequ
17	211.5	4.8	1025	23	ABR75711	Human poly(ADP-rib
18	204.5	4.6	168	20	AAH66319	Kidney injury asso
19	203	4.6	363	22	ABR94552	Human protein sequ
20	201	4.5	187	21	ABR41639	Human ORFX ORF1403
21	197	4.4	368	21	ABR43084	Human ORFX ORF2848
22	196.5	4.4	200	21	AAV71082	Arabidopsis thalia
23	196.5	4.4	239	21	AAV71082	Arabidopsis thalia
24	189	4.3	226	21	AAV58613	Human protein sequ
25	189	4.3	357	22	ABR78575	Human protein sequ
26	189	4.3	386	22	ABR11992	Human secreted pro
27	189	4.3	386	22	AAV79380	Human protein sequ
28	187.5	4.2	200	22	AAE01226	Human gene 13 enco
29	187	4.2	542	22	AAV71082	Human bone marrow
30	177	4.0	542	22	AAV71082	Human polypeptide
31	161.5	3.6	372	22	AAV54000	Human histone sequ
32	159	3.6	373	20	AAV50122	Human histone fusi
33	159	3.6	373	22	ABR30626	Amino acid sequenc
34	157.5	3.5	1181	22	ABR60894	Drosophila melanog
35	157.5	3.5	1181	22	ABR66297	Drosophila tankyria
36	157	3.5	378	21	ABR43763	Human cancer assoc
37	156	3.5	220	21	ABR50998	Human TRFA polype
38	156	3.5	487	22	AAV39115	Human polypeptide
39	154.5	3.5	1166	22	ABR47022	Human SPANK. Homo
40	151	3.4	173	22	AAV73955	Human colon cancer
41	148	3.3	125	22	AAV81296	Human AFP protein
42	148	3.3	716	22	AAU36810	Staphylococcus aur
43	146.5	3.3	1099	22	ABR66301	Human breast cancer
44	145	3.3	907	22	ABR48574	Human breast cancer
45	144	3.2	151	21	AAV21738	Arabidopsis thalia

ALIGNMENTS

RESULT 1	AAV71082	AAV71082 standard; Protein; 854 AA.
ID	AAV71082	
XX	AAV71082;	
AC	29-AUG-2000 (first entry)	
XX		
DT	Human B-aggressive lymphoma (BAL) protein.	
XX		
DE	Human; B-aggressive lymphoma; BAL; chromosome 3q21; tumor; malignancy;	
KW	differential expression; DLB-CL; Diffuse large B-cell lymphoma;	
KW	cytostatic; vaccine; treatment; diagnosis; non-Hodgkin's lymphoma;	
KW	cellular adhesion; sarcoma; carcinoma; myeloma.	
XX		
OS	Homo sapiens.	
XX		
Key	Location/Qualifiers	
FT	17..51	/note="Alternatively spliced sequence"
FT	136..256	/note="contains a duplicated domain of unknown function
FT		which is found in the non-histone region of histone
FT		macro H2A, non-structural polypeptides of sRNA viruses
FT		or in a family of related proteins from bacteria to
FT		eucaryotes"
FT	335..447	/note="contains a duplicated domain of unknown function
FT		which is found in the non-histone region of histone
FT		macro H2A, non-structural polypeptides of sRNA viruses
FT		or in a family of related proteins from bacteria to
FT		eucaryotes"
FT	160..288	
FT	Region	

```

FT FT /note="This region is 22% identical to histone
FT FT macro-H2A.1 protein"
FT FT 319..485
FT FT /note="This region is 26% identical to histone
FT FT macro-H2A.1 protein"
FT FT 392..395
FT FT /note="Tyrosine phosphorylation site"
FT FT 495..498
FT FT /note="Tyrosine phosphorylation site"
FT FT 508..709
FT FT /note="Partially homologous to protein families,
FT FT myosin heavy chain and cytoskeleton linkers
FT FT ezrin-radixin-moesin (ERM)"
FT FT 781..786
FT FT /label=Proline_rich_domain
XX XX WO200026331-A1.
XX XX 11-MAY-2000.
XX XX PD
XX XX 29-OCT-1999; 99WO-US25439.
XX XX PF
XX XX 29-OCT-1998; 98US-0106383.
XX XX PR
XX XX 30-OCT-1998; 98US-0106448.
XX XX PA
XX XX (DAND ) DANA FARBER CANCER INST INC.
XX XX
XX XX Shipp M, Aguilar R, Yakushijiin Y;
XX XX
XX XX WPI: 2000-365563/31.
XX XX DR
XX XX N-PSDB: AAD00629.
XX XX
XX XX New B-aggressive lymphoma nucleic acid for identifying cells exhibiting
XX XX or predisposed to malignancies such as lymphoma, sarcoma, carcinoma and
XX XX myeloma
XX XX
XX XX Claim 12; Fig 1; 151pp; English.
XX XX
XX XX The present sequence is a B-aggressive lymphoma (BAL).
XX XX protein obtained by screening 20 human cDNA libraries derived from
XX XX anti-immunoglobulin activated splenocytes and the Raji Burkitts lymphoma
XX XX cell line cloned into pCDM8. The BAL locus is mapped to chromosome 3q21.
XX XX BAL was identified to be differentially expressed in DLB-CL (Diffuse
XX XX large B-cell lymphoma), the most common non-Hodgkin's lymphoma, by
XX XX differential display technique. It was found to be more abundant in
XX XX tumours from patients with high risk fatal DLB-CL disease than low risk
XX XX cured disease. BAL transcripts are most abundant in lymphoid organs e.g.
XX XX spleen, foetal liver and peripheral blood, and several non-haematopoietic
XX XX organs e.g. heart and skeletal muscle. BAL is involved in modulation of
XX XX cellular adhesion and aggressiveness/severity of malignancy such as
XX XX DLB-CL. BAL nucleic acid and protein are useful for identifying cells
XX XX exhibiting or predisposed to malignancies such as lymphoma, sarcoma,
XX XX carcinoma and myeloma. BAL proteins are useful as immunogens to raise
XX XX anti-BAL antibodies. BAL modulators or BAL molecules can be used for
XX XX prophylactic and therapeutic treatment of a subject susceptible to or
XX XX having a disorder associated with aberrant BAL expression or activity,
XX XX such as non-Hodgkin's lymphoma. It is also useful in vaccine preparation.
XX XX
XX XX Sequence 854 AA:
XX XX
XX XX Query Match 100.0%; Score 4447; DB 21; Length 854;
XX XX Best Local Similarity 100.0%; Pred. No. 0;
XX XX Matches 854; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX XX
OY 1 MDFSMVGAAGAAAYNKSGRITSLILEFOKVAQIFPPMKRGNTDECLPYKCGSETGALGENY 60
Db 1 MDFSMVGAAGAAAYNKSGRITSLILEFOKVAQIFPPMKRGNTDECLPYKCGSETGALGENY 60
OY 61 SWQIPIHNHDFILKNNEROLCEYLONKRCFCISTLVSPVQEGNSKSLQVFRKMLPPIREL 120
Db 61 SWQIPIHNHDFILKNNEROLCEYLONKRCFCISTLVSPVQEGNSKSLQVFRKMLPPIREL 120
OY 121 SVWKDQLTTHAVDAVYVNAEDLLHGGLALALVKGAFGEIOESKQFVARYGKVSAGEI 180

```

D	121	SVKKDDLTTHAVDAVVAANAEDLLHGGLLALALYKAGCFEIOESKQFVARYGVSAGET	180
Q	181	AVTGAGRLPCKOIIHAVCPHNMEDKOGCTGLORALIVLSILNVYIKNTHIKTVAIPALS	240
D	181	AVTGAGRLPCKOIIHAVGPRHMEWDKOGCTGLORALIVLSILNVYIKNTHIKTVAIPALS	240
Q	241	SGIQFPLNLCTKTIYETIVINSLOGKRPMNSLKIHLIVSNEDPRTVAAFKASEFIIKSE	3000
Q	301	LGQETPESFNAMVNNLTLOIVOGHIEQMOTADYIVNSVNHDTIVGVVASSILOQAGVEM	3600
D	301	LGQETPESFNAMVNNLTLOIVOGHIEQMOTADYIVNSVNHDTIVGVVASSILOQAGVEM	3600
Q	361	KSEFLATKAKOFQSOQLVLTGKFNLECKYIYHVLHSHSEPPKOLIKHAMKECLEKIEQ	4200
D	361	KSEFLATKAKOFQSOQLVLTGKFNLECKYIYHVLHSHSEPPKOLIKHAMKECLEKIEQ	4200
Q	421	NITSISPAIGTGMELIKKETAAELDEYLTFAKHVKHQLTVKFIPTDLEIYKAFS	4800
D	421	NITSISPAIGTGMELIKKETAAELDEYLTFAKHVKHQLTVKFIPTDLEIYKAFS	4800
Q	481	SEMAKRSKMLSLNNYVSPQSTREEKRENGLEARSDAINLNGFNVEEYEAHAWIQRILSL	5400
D	481	SEMAKRSKMLSLNNYVSPQSTREEKRENGLEARSDAINLNGFNVEEYEAHAWIQRILSL	5400
Q	541	QNHIIENNNHILYGRKEHDILSLOKTSVSTELTISPGTELEIEGARADILEVYMI	6000
D	541	QNHIIENNNHILYGRKEHDILSLOKTSVSTELTISPGTELEIEGARADILEVYMI	6000
Q	601	EDMLCKYQOEEMARKKEKGLMSLGOPTLOOKTODEKKNITFLKCPVPTQELLOOKO	6600
D	601	EDMLCKYQOEEMARKKEKGLMSLGOPTLOOKTODEKKNITFLKCPVPTQELLOOKO	6600
Q	661	FEKCGLOLVKXEKIDNEVULMAAFQRRKKRMEBEKLNROPVSHRLFOQVYVQFCNVYCVGF	7200
D	661	FEKCGLOLVKXEKIDNEVULMAAFQRRKKRMEBEKLNROPVSHRLFOQVYVQFCNVYCVGF	7200
Q	721	QRMSTPCDDPKYGGAGIYFTKMLKLAERAKKISADKLIVFEAEVLTGFCOGHPRLNIY	7800
D	721	QRMSTPCDDPKYGGAGIYFTKMLKLAERAKKISADKLIVFEAEVLTGFCOGHPRLNIY	7800
Q	781	PRPLSPGALIOHSDVNVNVSPEFFVLFSGQALPOLYLMTCQGYVOSDYSSGPMRFA	8400
D	781	PRPLSPGALIOHSDVNVNVSPEFFVLFSGQALPOLYLMTCQGYVOSDYSSGPMRFA	8400
Q	841	QHPMRGASGSPVD 854	
D	841	QHPMRGASGSPVD 854	
RESULT 2			
TD	ABB05392	standard; Protein: 819 AA.	
XX	ABB05392:		
AC	10-Apr-2002	(first entry)	
XX	Human interferon-induced tetraspan (IT) protein #2.		
DE	Human: interferon-induced tetraspan; IT; cytosolic; anti-HIV;		
XX	immunosuppressive; antiinflammatory gene therapy; autoimmune disease;		
KW	inflammatory disease; cardiovascular disease; metabolic disease;		
KW	cancer growth inhibition; metastasis.		
XX	Homo sapiens.		
OS			
XX	Key	Location/Qualifiers	
FH	Modified-site	24..26	
FT	/note-	"asn is N-glycosylated"	
FT	Modified-site	69..71	

FT /note= "Asn is N-glycosylated"
 FT 198..217
 FT /label= "TM"
 FT /note= "transmembrane domain"
 FT Region 218..269
 FT /note= "intracellular loop"
 FT Domain 270..291
 FT /label= "TM"
 FT /note= "transmembrane domain"
 FT Region 292..730
 FT /note= "large extracellular loop"
 FT Modified-site 386..388
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 459..461
 FT /note= "Asn is N-glycosylated"
 FT Domain 731..754
 FT /label= "TM"
 FT /note= "transmembrane domain"
 FT Region 755..768
 FT /note= "intracellular loop"
 FT Domain 769..788
 FT /label= "TM"
 FT /note= "transmembrane domain"

W0200196399-A2.

20-DEC-2001.

15-JUN-2001; 2001WO-US19305.

15-JUN-2000; 2000US-211565P.

(CURA-) CURAGEN CORP.

(BIO) BIOGEN INC.

Peyman JA, Da Silva A, Hockman P;

WPI: 2002-098056/13.

DR N-PSDB: ABA93022.

Novel isolated interferon-induced tetraspan polypeptide, useful for treating autoimmune, inflammatory, cardiovascular and metabolic diseases, and for inhibiting cancer growth and metastasis -

Claim 1; Fig 1; 105pp; English.

CC The present sequence represents a human interferon-induced tetraspan (ITP) protein (I). (I) has cytosolic, anti-HIV, immunosuppressive and anti-inflammatory activities. The polynucleotide (II) encoding (I) can be used in gene therapy. (I) and (II) can be used for treating or preventing a pathology associated with (I) in a subject, preferably human, or for the manufacture of a medicament for treating a syndrome associated with a human disease selected from a pathology associated with (I), (I) and (II) can be used for treating autoimmune diseases, inflammatory diseases, cardiovascular diseases, metabolic diseases, and for inhibiting cancer growth and metastasis. (I) or (II) can be used as diagnostic and/or prognostic markers, in gene therapy, as research tools, and for tissue regeneration in vitro and in vivo.

SQ Sequence 819 AA;

Query Match 94.6%; Score 4208.5; DB 23; Length 819;
 Best Local Similarity 95.4%; Pred. No. 0;
 Matches 815; Conservative 3; Mismatches 1; Indels 35; Gaps 1;

QY 1 MDESMVAGAAAYNKRSGITSLILPQKVFQIRPQMRKGTRECLPKCCEGTALGENTY 60
 DB 1 MDESMVAGAAAYNKR-----SETGALGENTY 25
 QY 61 SMQIPINHNDRKILKNRROLCEVLONKFGCISTLSPVQSGNSKSLQVFRKNLTPREL 120
 DB 26 SMQIPINHNDRKILKNRROLCEVLONKFGCISTLSPVQSGNSKSLQVFRKNLTPREL 85

QY 121 SWMKDILTTHAVDAVNNANEDLHGGGLALALYKAGGFEOESKOFVARYKVSAGEI 180
 DB 86 SWMKDILTTHAVDAVNNANEDLHGGGLALALYKAGGFEOESKOFVARYKVSAGEI 145
 QY 181 AVTGAGRLPCAKQIIHAVGPRMWMENDKQCTGKLOBAIVSLNVIYKTKHTKVAIPALS 240
 DB 146 AVTGAGRLPCAKQIIHAVGPRMWMENDKQCTGKLOBAIVSLNVIYKTKHTKVAIPALS 205
 QY 241 SGIFQFPLNLCTKTIIVETIRVSLQCKPMMSNIKEIHLVSNEDPYAAKASEFLGKSE 300
 DB 206 SGIFQFPLNLCTKTIIVETIRVSLQCKPMMSNIKEIHLVSNEDPYAAKASEFLGKSE 265
 QY 301 LGOETTPSFNNMNVNLLQIVQGHIEWQTADVIVNSNPNHDITGPAKSLIOOAGEM 360
 DB 266 LGOETTPSFNNMNVNLLQIVQGHIEWQTADVIVNSNPNHDITGPAKSLIOOAGEM 325
 QY 361 KSEFLATRAKQFQRSQVLVTKGFNLFCIKYIYHVLHSEFPKPOLIKHAMKECLEKCEIQ 420
 DB 326 KSEFLATRAKQFQRSQVLVTKGFNLFCIKYIYHVLHSEFPKPOLIKHAMKECLEKCEIQ 385
 QY 421 NITSTSPALGTGNNIEIKKETAELLPEDVLTFAKDHYKHQLYKVFVFPDLELYKAFS 480
 DB 386 NITSTSPALGTGNNIEIKKETAELLPEDVLTFAKDHYKHQLYKVFVFPDLELYKAFS 445
 QY 481 SEMAKRSKMLSLNNYSVPQSTREEKRENGLEARSAPAINLMGFNVBEMEFAHAMIORITSL 540
 DB 446 SEMAKRSKMLSLNNYSVPQSTREEKRENGLEARSAPAINLMGFNVBEMEFAHAMIORITSL 505
 QY 541 QNHIIENNHHILYGRKRDHLSQLOKTSVSITEIISPGRTLEIEGARADLIEVNNI 600
 DB 506 QNHIIENNHHILYGRKRDHLSQLOKTSVSITEIISPGRTLEIEGARADLIEVNNI 565
 QY 601 EDMLCCKVOEEMARKKERGLMSLGOWTIOQKTODEMKNIEIFLCPYPPOELIDOKKO 660
 DB 566 EDMLCCKVOEEMARKKERGLMSLGOWTIOQKTODEMKNIEIFLCPYPPOELIDOKKO 625
 QY 661 FEKCGLOVLKYEKIDNEVLMFAFORKKRMEBEKILHROPVSHRLFQOVYQFCNVVCRGTF 720
 DB 626 FEKCGLOVLKYEKIDNEVLMFAFORKKRMEBEKILHROPVSHRLFQOVYQFCNVVCRGTF 685
 QY 721 QRMVSTPCDDPKYGAGIYFTKLNKLAERAKKISADKLIYFEAVLTGFCQGHPLNIV 780
 DB 686 QRMVSTPCDDPKYGAGIYFTKLNKLAERAKKISADKLIYFEAVLTGFCQGHPLNIV 745
 QY 781 PPLSPGALDGHDSVDVNVSSPEFVIRFGMOAIPQVLYMTGCOEYVOSODYSNGPMRPEA 840
 DB 746 PPLSPGALDGHDSVDVNVSSPEFVIRFGMOAIPQVLYMTGCOEYVOSODYSNGPMRPEA 805
 QY 841 QHPWRGFAAGSPVD 854
 DB 806 QHPWRGFAAGSPVD 819

RESULT 3
 ABB05391 standard; Protein: 780 AA.
 ID ABB05391;
 AC ABB05391;
 XX 10-Apr-2002 (first entry)
 DT Human interferon-induced tetraspan (ITP) protein #1.
 XX Human: interferon-induced tetraspan; ITP; cytosolic; anti-HIV;
 DE immunosuppressive; anti-inflammatory gene therapy; autoimmune disease;
 KW inflammatory disease; cardiovascular disease; metabolic disease;
 KW cancer growth inhibition; metastasis.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FH Misc-difference 417..419
 FT /note= "encoded by ACATT"

FT Misc-difference 770..772
 /note="encoded by ACATT"
 XX
 XX WO200196399-A2.
 XX
 XX 20-DEC-2001.
 PD
 XX 15-JUN-2001; 2001WO-US19305.
 PE
 XX 15-JUN-2000; 2000US-211565P.
 PR
 XX
 XX (CURA-) CURAGEN CORP.
 PA (BIOJ) BIOGEN INC.
 XX
 P1 Peyman JA, Da Silva A, Hockman P;
 XX
 XX WPI: 2002-098056/13.
 DR N-PSDB: ABA93022.
 DR
 XX
 PT Novel isolated interferon-induced tetraspan polypeptide, useful for
 PT treating autoimmune, inflammatory, cardiovascular and metabolic
 PT diseases, and for inhibiting cancer growth and metastasis -
 PS
 XX
 PS Claim 1; Page 8; 105pp; English.
 CC The present sequence represents a human interferon-induced tetraspan
 CC (ITP) protein (I). (I) has cytostatic, anti-HIV, immunosuppressive and
 CC antiinflammatory activities. The polynucleotide (II) encoding (I) can be
 CC used in gene therapy. (I) and (II) can be used for treating or preventing
 CC a pathology associated with (I) in a subject, preferably human, or for
 CC the manufacture of a medicament for treating a syndrome associated with a
 CC human disease selected from a pathology associated with (I). (I) and (II)
 CC can be used for treating autoimmune diseases, inflammatory diseases,
 CC cardiovascular diseases, metabolic diseases, and for inhibiting cancer
 CC growth and metastasis. (I) or (II) can be used as diagnostic and/or
 CC prognostic markers, in gene therapy, as research tools, and for tissue
 CC regeneration in vitro and in vivo.
 CC
 SQ Sequence 780 AA:
 Query Match 88.3%; Score 3927.5; DB 23; Length 780;
 Best Local Similarity 94.5%; Pred. No. 0;
 Matches 770; Conservative 3; Mismatches 5; Indels 37; Gaps 3;
 QY 1 MDFSVMAGAAAYNEKSGRTSLSLFQKVFQAQIFPQMRKGTRECLPKCSEGTALGENTY 60
 DB 1 MDFSVMAGAAAYNEK-----SETGALGENTY 25
 QY 61 SMQIPINHNDRKILKNRQCEVLQNKFGCISTLVSPVQSGNSKLVFRRKMLTPRIEL 120
 DB 26 SMQIPINHNDRKILKNRQCEVLQNKFGCISTLVSPVQSGNSKLVFRRKMLTPRIEL 85
 QY 121 SVMKDDLTTHAVDAVNVANANEDLLHGGGLALALYKAGFETQESKOPVARYGVSAGEI 180
 DB 86 SVMKDDLTTHAVDAVNVANANEDLLHGGGLALALYKAGFETQESKOPVARYGVSAGEI 145
 QY 181 AVTAGAGRLPCQKQIIHANGPRMMEWDKQCTGLQRAIVSILNVYIYKNTHTKVAIPALS 240
 DB 146 AVTAGAGRLPCQKQIIHANGPRMMEWDKQCTGLQRAIVSILNVYIYKNTHTKVAIPALS 205
 QY 241 SCGIOFPLNICTKTIVETIRVSLQKPMMSMLKEIHLVSNDDPVYAAFKASEFTLGSE 300
 DB 206 SCGIOFPLNICTKTIVETIRVSLQKPMMSMLKEIHLVSNDDPVYAAFKASEFTLGSE 265
 QY 301 LGOETTSFNAMVNNLTLOLVQGHIEWQADIVNSVNPDIIVGPAKSILOQAGEM 360
 DB 266 LGOETTSFNAMVNNLTLOLVQGHIEWQADIVNSVNPDIIVGPAKSILOQAGEM 325
 QY 361 KSEFLATKAKQFORSQVLVTKGFNLFCKYLYHVLHMSFPKPOLKHAMKECLEKTEQ 420
 DB 326 KSEFLATKAKQFORSQVLVTKGFNLFCKYLYHVLHMSFPKPOLKHAMKECLEKTEQ 385
 QY 421 NITSISFALGTGMMEIKKETAAELIDEVLTAF-AKHVHKQLIVKFIPTDLEITKAF 479

DB 386 NITSISFALGTGMMEIKKETAAELIDEVLTAF-AKHVHKQLIVKFIPTDLEITKAF 445
 QY 480 SSEMAKSKMLSLNNYSVPOSTREKRENGLEARSPAINLMGFVNEVEYEAHAMIOHLS 539
 DB 446 SSEMAKSKMLSLNNYSVPOSTREKRENGLEARSPAINLMGFVNEVEYEAHAMIOHLS 505
 QY 540 LQNHIIENNHILYLGKREHILSLOLQTSVSTETIISPORTLEIEGARADILEVVMN 599
 DB 506 LQNHIIENNHILYLGKREHILSLOLQTSVSTETIISPORTLEIEGARADILEVVMN 565
 QY 600 IEDMLCKVQEEEMARKKRGRLRSLSQWTOOQKQODEKKEIIFLKCVPPTQELDQKK 659
 DB 566 IEDMLCKVQEEEMARKKRGRLRSLSQWTOOQKQODEKKEIIFLKCVPPTQELDQKK 625
 QY 660 QFEKCGLOVLKVERKIDNEVLMAAFQRRKKMEKRLHROPVSHRLFOQVPOFCVWVRVG 719
 DB 626 QFEKCGLOVLKVERKIDNEVLMAAFQRRKKMEKRLHROPVSHRLFOQVPOFCVWVRVG 685
 QY 720 FQRMVSTPCDPRKYGAGIYFTKLNKLAERAKKISADKLIYFAEVLTFGFCGHPINI 779
 DB 686 FQRMVSTPCDPRKYGAGIYFTKLNKLAERAKKISADKLIYFAEVLTFGFCGHPINI 745
 QY 780 VPPLSPGALDGHDSVVDNVSSEPTF-VIFSGMA 813
 DB 746 VPPLSPGALDGHDSVVDNVSSEPTIYIFSGMA 780
 RESULT 4
 ABB50253
 ID ABB50253 standard; Protein: 653 AA.
 AC ABB50253;
 DT 05-FEB-2002 (first entry)
 XX
 DE Human transcription factor TRFX-104.
 KW Human; transcription factor; TRFX; cell proliferative disease;
 KW autoimmune disease; inflammation; neurological disease;
 KW developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;
 KW neuroprotective; antiinflammatory; gene therapy.
 OS Homo sapiens.
 XX
 XX WO200172777-A2.
 PN
 XX
 PD 04-OCT-2001.
 XX
 XX 13-MAR-2001; 2001WO-US08117.
 PE
 XX
 XX 13-MAR-2000; 2000US-0188986.
 PR
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 XX Hillman JL, Baughn MR, Yue H, Lal P, Lu DM, Patterson C;
 PI Azimzai Y, Bandman O, Tang YT, Mathur P, Shah P, Au-Young J;
 PI Reddy R;
 XX
 XX WPI: 2001-570896/64.
 DR N-PSDB: ABA83077.
 PT Novel transcription factor polypeptides, used to treat diseases
 PT associated with altered activity and expression of TRFX, and to screen
 PT for agents capable of modulating its activity -
 PS
 PS Claim 1; Pages 247-249; 327pp; English.
 CC The present sequence is the protein sequence for a human transcription
 CC factor. The transcription factor and its coding sequence are useful in
 CC the diagnosis, treatment and prevention of diseases associated with
 CC altered expression of the transcription factor e.g. cell proliferative,
 CC autoimmune/inflammatory, neurological and developmental disorders. A

CC number of specific disorders/diseases are given in the specification,
 CC including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,
 CC allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic
 CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,
 CC Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,
 CC psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative
 CC colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's
 CC disease, stroke, and viral, bacterial, fungal and protozoal infections.

XX Sequence 653 AA;

Query Match 76.2%; Score 3388; DB 22; Length 653;
 Best Local Similarity 99.8%; Pred. No. 1.2e-278;
 Matches 652; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 202 MEMKOGCTGLOKRAIVSLNVIYKNTHTKTAIPALSSGIFOPPLNLCRTIVETRV 261
 DB 1 MEMKOGCTGLOKRAIVSLNVIYKNTHTKTAIPALSSGIFOPPLNLCRTIVETRV 60
 QY 262 SLQGRPMKSNLKEIHLVSNEDPTVAAPKASEPILGKSELQETTPSFNAVNNLTLOI 321
 DB 61 SLQGRPMKSNLKEIHLVSNEDPTVAAPKASEPILGKSELQETTPSFNAVNNLTLOI 120
 QY 322 VQGIEMQTAIVYNSVPHDITVGPVAKSLIQAGVEMKSEFLATKAKQORSOLVLT 381
 DB 121 VQGIEMQTAIVYNSVPHDITVGPVAKSLIQAGVEMKSEFLATKAKQORSOLVLT 180
 QY 382 KGFNLFCYIYVLMHSEFPKQILKHAMCKLEKICIONTISTSEPALGCGNNEIKKET 441
 DB 181 KGFNLFCYIYVLMHSEFPKQILKHAMCKLEKICIONTISTSEPALGCGNNEIKKET 240
 QY 442 AAEILFDEVLTFARDVHKHQLTVKRVFPTDLEIYKAFSSPMARKSKMLSNNTSVPOST 501
 DB 241 AAEILFDEVLTFARDVHKHQLTVKRVFPTDLEIYKAFSSPMARKSKMLSNNTSVPOST 300
 QY 502 REERENGLEKRSAPAINMGFVNEEMFAHMIQRIISLQNHHTIENNHTIYLGRKEDI 561
 DB 301 REERENGLEKRSAPAINMGFVNEEMFAHMIQRIISLQNHHTIENNHTIYLGRKEDI 360
 QY 562 LSOQKTSVSVITELISGRTLEIEGARADLIEVNMIEMLCKVQEMARKKERGLWR 621
 DB 361 LSOQKTSVSVITELISGRTLEIEGARADLIEVNMIEMLCKVQEMARKKERGLWR 420
 QY 622 SLGWTIQOQKTODEMKNITFLKCPVPPTQELLDOKKQFEGCGLVAKVEKIDNEVMA 681
 DB 421 SLGWTIQOQKTODEMKNITFLKCPVPPTQELLDOKKQFEGCGLVAKVEKIDNEVMA 480
 QY 682 AFORKKMEKELHROPVSHRLEFQOVYQFCNVVCRVGFQRMVSTPCDPKYGAGIYFTKN 741
 DB 481 AFORKKMEKELHROPVSHRLEFQOVYQFCNVVCRVGFQRMVSTPCDPKYGAGIYFTKN 540
 QY 742 LKNIAEKAKKISADKLIVFEAEVLNGFCQGHPLNTVPPPLSGALDGHDSVDVNS 801
 DB 541 LKNIAEKAKKISADKLIVFEAEVLNGFCQGHPLNTVPPPLSGALDGHDSVDVNS 600
 QY 802 PETEVIFSGMAIPOLYIMTCTQEVYVOSDYSSGPMRPAQHPMRGASGSPVD 854
 DB 601 PETEVIFSGMAIPOLYIMTCTQEVYVOSDYSSGPMRPAQHPMRGASGSPVD 653

RESULT 5
 AAY71083
 ID AAY71083 standard; Protein: 826 AA.

AC AAY71083;
 XX
 XX 29-AUG-2000 (first entry)
 DE Mouse B-aggressive lymphoma (BAL) protein-short form.
 XX
 KW Mouse: B-aggressive lymphoma; BAL; tumour; malignancy;
 KW differential expression; DLB-CL; Diffuse large B-cell lymphoma;
 KW cytosolic; vaccine; treatment; diagnosis; non-Hodgkin's lymphoma;

KW cellular adhesion; sarcoma; carcinoma; myeloma.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Domain 748..753
 FT /label= Proline-rich_domain
 FT Region 74..250
 FT /note="This region is 24% identical to histone
 FT macro-H2A.1 protein"
 FT 261..450
 FT /note="This region is 26% identical to histone
 FT macro-H2A.1 protein"

WO200026231-A1.
 PD 11-MAY-2000.
 XX
 PF 29-OCT-1999; 99W0-US25439.
 XX
 PR 29-OCT-1998; 98US-0106383.
 PR 30-OCT-1998; 98US-0106448.
 XX
 PA (DAND) DNA FARMER CANCER INST INC.
 PI Shipp M, Aguilar R, Yakushijin Y;
 XX
 DR WPI: 2000-365563/31.
 DR N-PSDB: AAD00630.
 XX
 PT New B-aggressive lymphoma nucleic acid for identifying cells exhibiting
 PT or predisposed to malignancies such as lymphoma, sarcoma, carcinoma and
 PT myeloma
 PT
 PS Claim 12; Fig 2: 15pp; English.

The present sequence is a short form of B-aggressive
 CC lymphoma (BAL) protein obtained from mouse (Balb-c) spleen.
 CC BAL was identified to be differentially expressed in DLB-CL (diffuse
 CC large B-cell lymphoma), the most common non-Hodgkin's lymphoma, by
 CC differential display technique. It was found to be more abundant in
 CC tumours from patients with high risk fatal DLB-CL disease than low risk
 CC cured disease. BAL transcripts are most abundant in lymphoid organs e.g.
 CC spleen, foetal liver and peripheral blood, and several non-haematopoietic
 CC organs e.g. heart and skeletal muscle. BAL is involved in modulation of
 CC cellular adhesion and aggressiveness/severity of malignancy such as
 CC DLB-CL. BAL nucleic acid and protein are useful for identifying cells
 CC exhibiting or predisposed to malignancies such as lymphoma, sarcoma,
 CC carcinoma and myeloma. BAL proteins are useful as immunogens to raise
 CC anti-BAL antibodies. BAL modulators or BAL molecules can be used for
 CC prophylactic and therapeutic treatment of a subject susceptible to or
 CC having a disorder associated with aberrant BAL expression or activity,
 CC such as non-Hodgkin's lymphoma. It is also useful in vaccine preparation.

XX
 XX Sequence 826 AA;

Query Match 55.2%; Score 2453; DB 21; Length 826;
 Best Local Similarity 61.4%; Pred. No. 4.5e-199;
 Matches 508; Conservative 96; Mismatches 206; Indels 18; Gaps 7;

QY 41 NTECLPYKCSFTGALGYNYSWQIPINHNDFKILKNRROLCFVLFNFGCISTLVSPVQ 100
 DB 2 DTWMAAPARERANNLSLEHYRQPIKINVFELIKSNSQICEVQKNKFCGISTLVSCVL 61
 QY 101 EGNKSLO-VERKMLTPRIELSVWKDILTTAAVDAVANAEDLHGGLALALVYKAGCF 159
 DB 62 AGSSSPAORVFRRTLIPIELSVWKDILTRHVYDAVANAENLHGSGLAGSLVKTGCF 121
 QY 160 EIOESKOFVARVYKVSAGELAVYAGARLPCKQIIHAYGPRMMDKCGCTGKIORATVS 219
 DB 122 EIOESKRIIANVGIISVGGIVITGAGRPLCHLIIHAAGPRVTVYNSOTAIELKFAIRN 181
 QY 220 ILNVIYKNTHTKTAIPALSSGIFOPPLNLCRTIVETIRVSLQGRPMKSNLKEIHLVS 279

```

182 ILDVYTVYDKIRKIVAIIPALSSGIFQEPPLDCTSIILETIRLKYFQDKQMEGNLKEIHLVS 241
QY 280 NEDPTVAAPKASEPILIGK--SELGOETTP--SFNMAVNNVLLQIVOGHIEKQTDVLI 334
Db 242 NEDPTVASPKSASSIILGRDLSSWGGPETDPASTMTLRIGRGLTLQIVOGCIEKQTDVLI 301
QY 335 VNSVNPDLITGVPAKSILOQAGVEMKSEFLATKAKOFORSQVLVTKGFNLFCFKYIYHV 394
Db 302 GNSGYMDQFKSGRAVQOSILRQAGVEMKEL--DKVNLSTDPQEVWVWIKGFKLSQCYVFNH 359
QY 395 LHWSEFPKPOLIKHAKKCELEKTEONITSPALGTGNNETKEKTAELTPEVLTFA 454
Db 360 AMHSQINKYQTLKDAKMSCKLEKLPDINSISFALGTGLDLKSTPAQIMFEVFAFA 419
QY 455 KDHYKHQTLKVFVIFPTDLEIYKAFSSSEMAKRSKMLSNYS--VPOSTREKRENGLE 511
Db 420 KEHEKELTKYKIVIFPVDEVEYKIFVYAEKMTKRSNELMLSGSGALALQWSSGDEORRGLE 479
QY 512 ARSPAINLMGFNVEEMYEAHAMTORILSLONHHIENNHIILYGRKEDHILISQLOKTSV 571
Db 480 AGSPAINLMGVKVGEMCEADQEMIERLLVSLDHIIENNHIILYLGKREHDVLESLQTSRV 539
QY 572 STEPLISPGRELEIEGARADLIEVNNIEMDLCKVOEEMARKERGLMRSLGQMTIQOQ 631
Db 540 SISSTVSPRATLEIKRPOADLDAVMKIECMCDVOEBAKREKMLMSLSGGTQOE 599
QY 632 KTDQEMKENIIFLKCPVPPTQELLDQKQKFEKGLQVLKVEKIDNEVLMAAFQKRRKME 691
Db 600 KL-DKMEESTYFORYPASLQELQDRKKQKFEKGLWVQVQEDINKYLLAFOKRRKME 658
QY 692 EKLHRQVSHRLFOQVYVRCNVYCRVGFQRMYSTPCDPKYGAGIYTKNKLNAEKAKK 751
Db 659 ERPRKGGSORLFOQVYHOCNCTVCRVGFHRMYSTSYNPPYAGIYTKNKLNAEDVKK 718
QY 752 ISADKLIYFEAEVLTCFQCGHPLNIVPPPLSPGALIDGDSVVDVNSPEFTVIRSGM 811
Db 719 TTSYDKLIYFEAEVLTCFQCGSNMIIIPPLSPGALDVNDVVDVNSPEFTVIRNGM 778
QY 812 QALPQYLTCTQOEY-----VQSDYSSGPMRPRAQHPRMGFASGSPV 853
Db 779 QAMPLYLMTCTQDRTFQSHPMMSQDYSFGPMVSSLSQSEWVLNGSSV 826

RESULT 6
AAM79380
ID AAM79380 standard; Protein: 360 AA.
AC AAM79380;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human protein SEQ ID NO 3026.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.

```

```

PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
XX N-PSDB; AAK52513.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy
XX
XX Claim 20; Page 233; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAK60302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activity/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM60020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.

SQ Sequence 360 AA;
QY 497 VPOSTREKRENGLEARSPAINLMGFNVEEMYEAHAMTORILSLONHHIENNHIILYGR 556
Db 3 VPOSTREKRENGLEARSPAINLMGFNVEEMYEAHAMTORILSLONHHIENNHIILYGR 62
QY 557 KEHDHLSQLOKTSVSTTEIISPGRELEIEGARADLIEVNNIEMDLCKVOEEMARKE 616
Db 63 KEHDHLSQLOKTSVSTTEIISPGRELEIEGARADLIEVNNIEMDLCKVOEEMARKE 122
QY 617 RGLMRSLGQMTIQOQKQODEMKENIIFLKCPVPPTQELLDQKQKFEKGLQVLKVEKIDN 676
Db 123 RGLMRSLGQMTIQOQKQODEMKENIIFLKCPVPPTQELLDQKQKFEKGLQVLKVEKIDN 182
QY 677 EVLMAAFQRRKKMEELKHQPVSHRLFOQVYVRCNVYCRVGFQRMYSTPCDPKYGAGI 736
Db 183 EVLMAAFQRRKKMEELKHQPVSHRLFOQVYVRCNVYCRVGFQRMYSTPCDPKYGAGI 242
QY 737 YFKNLKNLAEKAKKISAAKLIYFEAEVLTCFQCGHPLNIVPPPLSPGALIDGDSV 796
Db 243 YFKNLKNLAEKAKKISAAKLIYFEAEVLTCFQCGHPLNIVPPPLSPGALIDGDSV 302
QY 797 DNVSSPETVIFSGMQAIPQYLTCTQOEYVQSDYSSGPMRPRAQHPRMGFASGSPVD 854
Db 303 DNVSSPETVIFSGMQAIPQYLTCTQOEYVQSDYSSGPMRPRAQHPRMGFASGSPVD 360

RESULT 7
AAM78396
ID AAM78396 standard; Protein: 335 AA.
AC AAM78396;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human protein SEQ ID NO 1058.
XX

```


KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 DR WPI: 2001-476283/51.
 DR N-PSDB; AAK51529.
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 20; Page 3291; 6221pp; English.
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78333-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 CC
 XX
 SQ Sequence 335 AA:

Query Match 39.7%; Score 1765; DB 22; Length 335;
 Best Local Similarity 99.7%; Pred. No. 3.1e-141;
 Matches 334; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 520 MGFNVEEYEAHMTORILSLONHHIENNHILYGRKEHDILOLOKTSSTVITEIISP 579
 DB 1 MGFNVEEYEAHMTORILSLONHHIENNHILYGRKEHDILOLOKTSSTVITEIISP 60
 QY 580 GRTELEIEGARADILEVVMNIEDMLCKVOEEMARKKERGLMRSLOQWTIOOKTODEKKE 639
 DB 61 GRTELEIEGARADILEVVMNIEDMLCKVOEEMARKKERGLMRSLOQWTIOOKTODEKKE 120
 QY 640 NIILKCPVPPTOELLDOOKKOFKCGLOVLKVEKIDNEVLAARQKKKMEKILHROPV 699
 DB 121 NIILKCPVPPTOELLDOOKKOFKCGLOVLKVEKIDNEVLAARQKKKMEKILHROPV 180
 QY 700 SHRLFOOVYQFCNVVCGVGFQRMVSPDCPKYAGIYFTNKLNLAKAKKISADKLI 759
 DB 181 SHRLFOOVYQFCNVVCGVGFQRMVSPDCPKYAGIYFTNKLNLAKAKKISADKLI 240

QY 760 VYFAEVLITGFFCOGHPLNIVPPLSPALDGHDSVVDNVSSPETFVIFSGMAIPOYLW 819
 DB 241 VYFAEVLITGFFCOGHPLNIVPPLSPALDGHDSVVDNVSSPETFVIFSGMAIPOYLW 300
 QY 820 TCTOEYVOSDYSSGPMRPFQHPWRCGFASGSPVD 854
 DB 301 TCTOEYVOSDYSSGPMRPFQHPWRCGFASGSPVD 335
 RESULT 8
 ABP41342
 ID ABP41342 standard; Protein: 249 AA.
 XX
 AC ABP41342;
 XX
 DT 23-AUG-2002 (first entry)
 XX
 DE Human ovarian antigen HFXLL31, SEQ ID NO:2474.
 XX
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antineoplastic; gynaecological; reproductive.
 XX
 OS Homo sapiens.
 XX
 PN WO200200677-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 07-JUN-2001; 2001WO-US18569.
 XX
 PR 07-JUN-2000; 2000US-209467P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Birse CE, Rosen CA;
 DR WPI: 2002-147878/19.
 DR N-PSDB; ABQ54419.
 XX
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.
 PT ovarian cancer), immune disorders, cardiovascular disorders and
 PT neurological diseases -
 XX
 PS Claim 11; SEQ ID NO 2474; 2922pp; English.
 XX
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP41328) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and

CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 249 AA:
Query Match 24.8%; Score 1104; DB 23; Length 249;
Best Local Similarity 92.9%; Pred. No. 2.8e-85;
Matches 222; Conservative 2; Mismatches 9; Indels 6; Gaps 2;

OY 157 GGEFIOESKOPFARVKGVSAGETAVTGAARLPCKOIIHAVGPRMMDKOGCTGKQRA 216
DB 1 GGEFIOESKOPFARVKGVSAGETAVTGAARLPCKOIIHAVGPRMMDKOGCTGKQRA 60
OY 217 IVSLTAVVIAKNTHIKTVATPALSGLFQFPPLNCTKTIVETIRVSLQCKPMMSNLKEIH 276
DB 61 IVSLTAVVIAKNTHIKTVATPALSGLFQFPPLNCTKTIVETIRVSLQCKPMMSNLKEIH 120
OY 277 LVSNEDPTVAAFKASFEILGKSELGOETTPSFNANVNNLTQIVOGHIEMOTADIVN 336
DB 121 LVSNEDPTVAAFKASFEILGKSELGOETTPSFNANVNNLTQIVOGHIEMOTADIVN 180
OY 337 SVNPHDITVGPVAKSILOAGVEKSEFLATKAKOFOF-----SOLVLTGKGFLECKY 390
DB 181 SVNPHDITVGPVAKSILOAGVEKSEFLATKAKOFOF-----SOLVLTGKGFLECKY 238

RESULT 9
AAB93319
ID AAB93319 standard: Protein; 556 AA.
XX
AC AAB93319;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12406.
XX
KM Human; primer: detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 12406; 2537pp + CD ROM; English.

XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-AT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX
SQ Sequence 556 AA:
Query Match 9.4%; Score 420; DB 22; Length 556;
Best Local Similarity 26.6%; Pred. No. 1.2e-26;
Matches 144; Conservative 118; Mismatches 220; Indels 60; Gaps 22;

OY 111 RKLTPRIELSVKDDLTITAVAVVANAEDL-LHGGGLALALVKGFEIOESKOPF 169
DB 4 RLLRP-LQMLVKGEGQNAKTVDVNSVPLDLVLSGSPKSLLEAGELDEE-LDTV 61
OY 170 ARYKVSAGETAVTGAARLPCKOIIHAVGPRMMDKOGCTGKQRAIVSLN-VYIK 227
DB 62 GGGVAVSMGTVALKTSSWNLDRCYLVHAVPAEW-----RNGSTSSL-KIMEDIHECMITE 116
OY 228 NTHIKTVATPALSGLFQFPPLNCTKTIVETIRVSLQCKPMMSNLKEIH-LVSNEDPTV 285
DB 117 SLSLKSTAPPAIGTCNIGFPKNTFAELISEV-FKFSKMQLTLEVHFLIPSDHEHT 175
OY 286 AAF-----KAASEFLGKSELGOETTPSFNA-----WVNNLTQIVOGHIEMO 329
DB 176 QAFSDEPARRANGNLVSDKIPKAKDTGFGYVSSPDGVEYENKIGSITITQVASGDTTKG 225
OY 330 TADIVNSV-NPHDITVGPVAKSILOAGVEKSEFLATKAKOFOF-----SOLVLTGKGFLECKY 388
DB 236 EADIVNSTNSNFKLKG-VSKALIECAGQNVERE---CSQAOQQRNDIITGGGFLRC 291
OY 389 KYIVHVLHSEFPKPOLIKHAMKECLEKIEONITISIPALGTGNNETKETAETLFD 448
DB 292 KNIHVIHGND-----VKSSVSVLQECERKNVSSICLPAITGNAKQHPDXAEAID 345
OY 449 EVLTFE-KDHVKHOLTVKFIIF-PTDEITKA-----ESSEAKRKRKMSLNYSV 457
DB 346 AIEDFVQKSAQSVKRVVIFLPQYLDVEYANMKKREGTQLSSQOSVMSKLSFLGFS- 404
OY 498 POSTREKRENGLEAR--SPAINLMGFVDEMYEAHMIOIRIISL-ONHHIIRNNHLLYL 554
DB 405 -KOSPQKKHNLVLEKKTESATFRYCGENVCVEYATISWDLDLIEKDCPTTSEDECKDF 463
OY 555 GKREHDILOLOKTSVSIETIISPGTELEIGARADLIEVYNNIEMLCKYQVEEMARK 614
DB 464 DEKEYQELNELQKKLINIS--LDHKRPLIKVIGISDVQARDEIRAMIKRVR--LAKE 519
OY 615 KE 616
DB 520 QE 521

RESULT 10
AAB01176

PI Buchrieser C, Frangoul L, Couve E, Rusnick C, Esli H, Dehoux P,
 PI Dussurget O, Cheonani F, Nedjati H, Glaser P, Kunst F, Cossart P,
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA,
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,
 PI Chakraborty T, Domm E, Hain T, Berche P, Charbit A, Durant L,
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,
 PI Madueno E, de Pablos B, Wehlund J, Kaerst U, Entian K, Hauf J,
 PI Rose M, Voss H;
 XX WPI: 2002-010914/01.
 XX
 XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment
 PT and prevention of *Listeria* and related bacterial infections, and
 PT related polypeptides -
 XX
 XX Clalm 6; SEQ ID NO 2058; 192pp; French.
 XX
 CC The present invention relates to the genome sequence of *Listeria*
 CC *monocytogenes* ESD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in *L.*
 CC *monocytogenes* and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of *L. monocytogenes* and related organisms,
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate *L. monocytogenes*-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by *L.*
 CC *monocytogenes* and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 XX Sequence 176 AA;
 XX
 XX Query Match 5.3%; Score 235; DB 23; Length 176;
 XX Best Local Similarity 32.2%; Pred. No. 1.1e-11;
 XX Matches 55; Conservative 39; Mismatches 73; Indels 4; Gaps 3;
 XX
 OY 118 IELSYWKDLDLTHAVDAVANNANEDLLHGGGLALALVKGFEIODESKOFVARYKSYA 177
 DB 1 MEITIVKGDITPEQEVDAIVNANNSGLLGCGVDGAIHQAGGDLKEQCEVYNRIGSCPA 60
 OY 178 GEIANTGAGRLPCKOITHAIVGPRMWMKQCGCTKLRRAIVSILNVIYKNTHTKTAIP 237
 DB 61 GEAVTTSAGDLKATYTIHAVGPIMKDGHOE-ANKLASCYWKALDLAGKCD--LTSIAFP 117
 OY 238 ALLSGIFQFPLNLCTKTIVETIRVSLQCKPMNSLKEIHLVSNEDPTVAAP 288
 DB 118 NISTGVTPPKLALEVALYTVRKMAE-EYDTSIKETIRFVCEDEENLKY 167
 XX
 XX RESULT 15
 XX AAU58306
 XX ID AAU58306 standard; Protein: 246 AA.
 XX
 XX AAU58306;
 XX
 XX 27-FEB-2002 (first entry)
 XX
 DE Propionibacterium acnes immunogenic protein #19202.
 XX
 XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopthic; neuroprotectant.
 XX
 XX Propionibacterium acnes.
 XX
 XX WO200181581-A2.

XX
 PD 01-NOV-2001.
 XX
 XX 20-APR-2001; 2001WO-US12865.
 XX
 XX 21-APR-2000; 2000US-199047P.
 PR 02-JUN-2000; 2000US-208841P.
 PR 07-JUL-2000; 2000US-216747P.
 XX
 XX (CORI) CORIXA CORP.
 XX
 XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L.maisonneuve J, Zhang Y, Jen S, Carter D;
 XX
 XX WPI: 2001-616774/71.
 DR N-PSDB: AAS59590.
 XX
 XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX
 XX Example 1; SEQ ID NO 19501; 1069pp; English.
 XX
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC *P. acnes* is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of *P. acnes* in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for *P. acnes* proteins. These antibodies can be used to
 CC downregulate expression and activity of *P. acnes* polypeptides and
 CC therefore treat *P. acnes* infections. The antibodies may also be used as
 CC diagnostic agents for determining *P. acnes* presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 XX Sequence 246 AA;
 XX
 XX Query Match 4.9%; Score 219; DB 22; Length 246;
 XX Best Local Similarity 27.2%; Pred. No. 4.1e-10;
 XX Matches 61; Conservative 46; Mismatches 89; Indels 28; Gaps 6;
 XX
 OY 84 VLQKFCISTSLVSEVDEGNSKSLQVFRKMLTPRIELSVWKDLDLTHAVDAVANNANEDL 143
 DB 48 LLAASRSIAVTPDPSAEGDGRSCTI-----GAMADITLLRADITLTDVDAVANNANQL 102
 OY 144 LHGGGLALALVAKAGFEIODESKOF--VARYKVSAGIANTGAGRLPCKOITHAIVGPRW 201
 DB 103 AGGGGVDGAIHRAAGPELSQACRKLRRETTLDG-PTGQSVATFGKMPAKVHIPTVCPW 162
 OY 202 MEW-KDQCGCTKLRRAIVSILNVIYKNTHTKTAIPALSSGIFQFPLNLCTKTIVETIR 260
 DB 163 AKTIDKSDQLASCYRTSLHVADEI-----GARTIAFPITSAGCYGPMDEATRAVENC 217
 OY 261 VSLQCKPMNSLKEIHLVSNEDPTVAAPKASFTILGKSELGQ 304
 DB 218 QT-----VTKVDITLYLVAFAEAGYKRA-----LGE 246
 XX
 XX Search completed: May 13, 2003, 11:42:12
 XX Job time : 86 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 11:40:47 ; Search time 30 Seconds
(without alignments)
837.573 Million cell updates/sec

Title: US-09-830-762-2

Perfect score: 4447
Sequence: 1 MDFSVMAGAAAYNEKSGRIT.....PMRPQHPHWRGFSAGSPVD 854

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/protodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/protodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/protodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/protodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/protodata/1/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/protodata/1/1aa/Backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	165.5	3.7	368	2	US-08-824-878-3 Sequence 3, App11
2	165.5	3.7	368	4	US-09-353-688-3 Sequence 3, App11
3	159	3.6	373	2	US-08-824-878-1 Sequence 1, App11
4	159	3.6	373	4	US-09-353-688-1 Sequence 1, App11
5	136.5	3.1	1375	4	US-09-722-139-2 Sequence 2, App11
6	136.5	3.1	1375	4	US-09-721-832-2 Sequence 2, App11
7	136.5	3.1	1375	4	US-09-721-689-2 Sequence 2, App11
8	134	3.0	1166	4	US-09-350-982C-5 Sequence 5, App11
9	133	3.0	1786	4	US-08-973-462-8 Sequence 8, App11
10	132.5	3.0	2512	2	US-08-801-263A-9 Sequence 9, App11
11	132.5	3.0	2512	4	US-09-102-248-9 Sequence 9, App11
12	132	3.0	1327	4	US-09-102-387-2 Sequence 2, App11
13	131	2.9	2517	2	US-08-801-263A-5 Sequence 5, App11
14	131	2.9	2517	3	US-09-102-248-5 Sequence 5, App11
15	130.5	2.9	712	2	US-08-468-576B-17 Sequence 17, App11
16	130.5	2.9	712	3	US-08-468-577B-17 Sequence 17, App11
17	130.5	2.9	712	3	US-08-468-577B-17 Sequence 17, App11
18	130.5	2.9	2500	3	US-08-801-263A-2 Sequence 2, App11
19	130.5	2.9	2500	3	US-09-102-248-2 Sequence 2, App11
20	118	2.7	498	4	US-08-858-207A-310 Sequence 310, App11
21	118	2.7	2860	2	US-08-826-267-2 Sequence 2, App11
22	116	2.6	2037	4	US-09-306-998-3 Sequence 3, App11
23	115.5	2.6	630	4	US-08-973-462-9 Sequence 9, App11
24	114.5	2.6	1013	4	US-09-255-829-18 Sequence 18, App11
25	114	2.6	397	6	5187089-2 Patent No. 5187089
26	114	2.6	397	6	5187089-9 Patent No. 5187089
27	114	2.6	397	6	5187089-10 Patent No. 5187089

28	114	2.6	397	6	5187089-11 Patent No. 5187089
29	114	2.6	1053	4	US-09-724-519-2 Sequence 2, App11
30	114	2.6	1053	4	US-09-592-037-2 Sequence 2, App11
31	112	2.5	1038	4	US-09-541-782-4 Sequence 4, App11
32	112	2.5	1038	4	US-09-723-820-4 Sequence 4, App11
33	112	2.5	1388	4	US-09-572-191-2 Sequence 2, App11
34	112	2.5	1388	4	US-09-723-262-2 Sequence 2, App11
35	112	2.5	1388	4	US-09-723-219-2 Sequence 2, App11
36	111.5	2.5	514	2	US-08-960-023-14 Sequence 14, App11
37	111.5	2.5	550	2	US-08-844-058-2 Sequence 2, App11
38	111.5	2.5	3696	4	US-09-134-001C-5080 Sequence 5080, App11
39	111	2.5	397	6	5457090-2 Patent No. 5457090
40	111	2.5	397	6	5457090-4 Patent No. 5457090
41	111	2.5	397	6	5495001-7 Patent No. 5495001
42	110.5	2.5	1589	3	US-08-755-587-189 Sequence 189, App11
43	110	2.5	3248	1	US-08-353-700-1 Sequence 1, App11
44	110	2.5	3248	5	PCT-US85-16216-1 Patent No. 5187089
45	109.5	2.5	398	6	5187089-4 Patent No. 5187089

ALIGNMENTS

RESULT 1
US-08-824-878-3
Sequence 3, Application US/08824878
Patent No. 5981221
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: NOVEL HISTONE FUSION PROT
TITLE OF INVENTION: E1N
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/824, 878
FILING DATE: Filed Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0255 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 368 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 205276
US-08-824-878-3

Query Match 3.7% Score 165.5; DB 2; Length 368;
Best Local Similarity 24.1% Pred. No. 8e-07;
Matches 76; Conservative 47; Mismatches 140; Indels 53; Gaps 10;

[illegible]

RESULT 2
US-09-353-688-3
; Sequence 3, Application US/09353688

GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: NOVEL HISTONE FUSION PROT
TITLE OF INVENTION: E1N
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/353,688
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/824,878
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0255 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 368 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 205276
US-09-353-688-3

Query Match	3.7%;	Score 165.5;	DB 4;	Length 368
Best Local Similarity	24.1%;	Pred. NO. 8e-07;		

[illegible]

RESULT 3

; Sequence 1, Application US/08824878

```

1 GENERAL INFORMATION:
2 APPLICANT: Hillman, Jennifer L.
3 TITLE OF INVENTION: NOVEL HISTONE FUSION PROT
4 TITLE OF INVENTION: EIN
5 NUMBER OF SEQUENCES: 4
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: Incyte Pharmaceuticals, Inc.
8 STREET: 3174 Porter Drive
9 CITY: Palo Alto
10 STATE: CA
11 COUNTRY: USA
12 ZIP: 94304
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Diskette
15 COMPUTER: IBM Compatible
16 OPERATING SYSTEM: DOS
17 SOFTWARE: FastSeq for Windows Version 2.0
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/824,878
20 FILING DATE: Filed Herewith
21 CLASSIFICATION: 435
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER:
24 FILING DATE:
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Billings, Lucy J.
27 REGISTRATION NUMBER: 36,749
28 REFERENCE/DOCKET NUMBER: PF-0255 US
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 415-855-0555
31 TELEFAX: 415-845-4166
32 INFORMATION FOR SEQ ID NO: 1:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 373 amino acids
35 TYPE: amino acid
36 STRANDEDNESS: single
37 TOPOLOGY: linear
38 IMMEDIATE SOURCE:
39 LIBRARY: BRSNOT05
40 CLONE: 2297753
41 US-08-824-878-1

```

US-08-824-878-1

[illegible]

```

      5
US-09-722-139-2
: Sequence 2, Application US/09722139
: Patent No. 6355471
: GENERAL INFORMATION:
: APPLICANT: Beraud, Christophe
: APPLICANT: Freedman, Richard
: TITLE OF INVENTION: No. 6355471el motor proteins and methods for
: TITLE OF INVENTION: their use
: FILE REFERENCE: 1055
: CURRENT APPLICATION NUMBER: US/09/722.139
: CURRENT FILING DATE: 2000-11-24
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 1375
: TYPE: PRT
: ORGANISM: Human
US-09-722-139-2

Query Match      3.1%; Score 136.5; DB 4; Length 1375;
Best Local Similarity 18.7%; Pred. No. 0.0033;
Matches 154; Conservative 131; Mismatches 299; Indels 239; Gaps 38.

      2 DFSWAGAAANNEKGRITSLSLFQKYFAQIFPQWRKGNTECLPYKCEMGALGENYS 61
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 DSEFYSA----DTKSPDYVSQEMVFKPLGTVDVKSAFEGYNACVFAY-----GQTGSGKS 109

      62 WOPIPNHDFKLKNNEROLCEVLNKF-----GCISTLSPVPOEGNSKSLQFRKM 113
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 110 YTMKNSGDSGLIP----RICGGLSRINETTRMWDASRTVSYLEIYENVRDLRRK 165

      114 LTRPRLTSV-----WKDDLTTHAVDAVVAANEDLLHGGGLALALVKGAGFETQES 165
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 166 SSKFTNLRYREHPKEGYPVEDLSKHLVONYGDV--EELMDADGINRTTAATGMNDVSSRS 223

      166 KQFVARVKVSAGETAVNGA---GRLPCKOI--IHAVGRPAME-WDKQGT-----CK 212
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 224 HAIF-----IKFTQAKFDSMPCEIVYSKHLVLDLAGSERADATGATGVRLEKGGN 274

      213 LQRAIV-----SILNVIYTKNT--HIKTVAIPLASSGIFQFPLNICT--- 252
      : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 275 INKSLVTLGAKKQVFPYRDSVLTWLLKDSLGCNSKTIIMATISPADVNGETLSTLRY 334
QY 233 -----KTIYETIRVSL--OGKPMNSLKEIHLVSN-----EDPTVAFAKASEFI 295
Db 335 ANRAKNIINKPTINEDANVKILREIRAIARLKTLLAOGNIALDSTALSMEEKLQON 394
QY 296 LGR-SELGOETTPSFNAN--VYNNLTLOIVOGHIEMOTADVIYNSVNHDTIVGPVAKSI 352
Db 395 EARVOELTKEMTNKMNENTONILKEOTLALRREGI-----GVLDSELPHLI----- 440
QY 353 LOAGVEMKSEFLATKAKOFORSOLVLTGKFNLFCKYIYVLMHSEF-----PKPOI 405
Db 441 ----GID--DDLSTGI-----LHYHKGQTYVGRDASTEDDI 474
QY 406 LKHAMKECLEKICIONT--TSISFPALGT-----GNMEIKKET-----AAELFDEVLPFAK 455
Db 475 VLHGLDLESEHCIFENIGTITLPLSGSOSVNGVOYVEATHLNOGAVILLGRTNMF 534
QY 456 DHVKHQLTVKRVFIPPTDEIYKAFSESEMAKRSKMLSNYSVPOSTREKREN----- 508
Db 535 NHPKE-----AAKLRKRKSGLLSSFSLSMTDLKSRENLSAVMLY 575
QY 509 --GLE-ARSPAINLMGFN-----VEEMEYEAH--AWIOTI-----LSLQ 541
Db 576 NPGLEFEHQOELELEKLSKRKLIEEMEKOKSDKAELERMQOEVEYORKETEIVOLQIR 635
QY 542 NHHIENNNHILYLRKREDILSOLQTSVSIITELISPGRTLEIEGARADLIEVNMIE 601
Db 636 KOESLKRSPHIENKLDLAIEKEKEFEERLRE-----QOEILOKKRQOE-----E 682
QY 602 DMLCKVOEMARKKRGILMSLQGTI-----OOQKODEKKENIIFLKCVPTPOELLD 656
Db 683 ETLFVRQOELORLKELNNEKAKEKFOIFQELDOLOKEKEDQYAKLELEK-----KRL 736
QY 657 OKKOFKCGLOVLVKEIDNEVLMAAFQRRKKMMEKHLRQPV 699
Db 737 OEKE-----QVMLVAHLEQL-----REKQEMIQLLRGEV 767

RESULT 6
US-09-721-832-2
: Sequence 2, Application US/09721832
: Patent No. 6399346
: GENERAL INFORMATION:
: APPLICANT: Beraud, Christophe
: APPLICANT: Freedman, Richard
: TITLE OF INVENTION: No. 6399346e1 motor proteins and methods for
: FILE REFERENCE: 1055
: CURRENT APPLICATION NUMBER: US/09/721,832
: CURRENT FILING DATE: 2000-11-24
: NUMBER OF SEQ. ID NOS: 4
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 1375
: TYPE: PR1
: ORGANISM: Human
US-09-721-832-2

Query Match 3.1%; Score 136.5; DB 4; Length 1375;
Best Local Similarity 18.7%; Pred. No. 0.0033;
Matches 154; Conservative 131; Mismatches 299; Indels 239; Gaps 38;

QY 2 DFSWAGAAVNEKSGRTISLSLFQVFAQIFPQWRKGNTECLPYKCSFTGALGENYS 61
Db 59 DFSFYSA-----DTKSPDYVSOEMVFKLTGTDVYKSAFEGYACVFAY-----GOTGSGKS 109
QY 62 WQIPIINDFKILKNNEKROCEVLONKF-----GCISLTVSPVQEGNSKSLQVFRKM 113
Db 110 YTMGNNGSDSLIP-----RICEGLFSKRNTETRWDEASFRTEVYLYLEYNKRVLDLRRK 165
QY 114 LTPRIELSV-----WKDULTTHAVDAVVAANEDLLHGGGLALALVKGGEIOEES 165

```

```

Db 166 SSKTFNLVRREHPKPEGYEDLSKHLVONYGDV--EELMDAGNINRTTAATGMDVSSRS 223
QY 166 KQFVARYKVSAGEIAYVGA---GRLPCKQI--IHAVGPRWME--WDKOGCT-----GK 212
Db 224 HAIF-----IKFQAKFDSMPCEYVSKIHVLVASSERDATTGATGVRLKEGN 274
QY 213 LQRAIV-----SILNVIYKNT--HIKTAIPALSSGIFQPLNCT--- 252
Db 275 INKSLVTLGAKKQVFPYRDSVLTWLLKDSLGCNSKTIIMATISPADVNGETLSTLRY 334
QY 253 -----KTIYETIRVSL--OGKPMNSLKEIHLVSN-----EDPTVAFAKASEFI 295
Db 335 ANRAKNIINKPTINEDANVKILREIRAIARLKTLLAOGNIALDSTALSMEEKLQON 394
QY 296 LGR-SELGOETTPSFNAN--VYNNLTLOIVOGHIEMOTADVIYNSVNHDTIVGPVAKSI 352
Db 395 EARVOELTKEMTNKMNENTONILKEOTLALRREGI-----GVLDSELPHLI----- 440
QY 353 LOAGVEMKSEFLATKAKOFORSOLVLTGKFNLFCKYIYVLMHSEF-----PKPOI 405
Db 441 ----GID--DDLSTGI-----LHYHKGQTYVGRDASTEDDI 474
QY 406 LKHAMKECLEKICIONT--TSISFPALGT-----GNMEIKKET-----AAELFDEVLPFAK 455
Db 475 VLHGLDLESEHCIFENIGTITLPLSGSOSVNGVOYVEATHLNOGAVILLGRTNMF 534
QY 456 DHVKHQLTVKRVFIPPTDEIYKAFSESEMAKRSKMLSNYSVPOSTREKREN----- 508
Db 535 NHPKE-----AAKLRKRKSGLLSSFSLSMTDLKSRENLSAVMLY 575
QY 509 --GLE-ARSPAINLMGFN-----VEEMEYEAH--AWIOTI-----LSLQ 541
Db 576 NPGLEFEHQOELELEKLSKRKLIEEMEKOKSDKAELERMQOEVEYORKETEIVOLQIR 635
QY 542 NHHIENNNHILYLRKREDILSOLQTSVSIITELISPGRTLEIEGARADLIEVNMIE 601
Db 636 KOESLKRSPHIENKLDLAIEKEKEFEERLRE-----QOEILOKKRQOE-----E 682
QY 602 DMLCKVOEMARKKRGILMSLQGTI-----OOQKODEKKENIIFLKCVPTPOELLD 656
Db 683 ETLFVRQOELORLKELNNEKAKEKFOIFQELDOLOKEKEDQYAKLELEK-----KRL 736
QY 657 OKKOFKCGLOVLVKEIDNEVLMAAFQRRKKMMEKHLRQPV 699
Db 737 OEKE-----QVMLVAHLEQL-----REKQEMIQLLRGEV 767

RESULT 7
US-09-721-689-2
: Sequence 2, Application US/09721689
: Patent No. 6440685
: GENERAL INFORMATION:
: APPLICANT: Beraud, Christophe
: APPLICANT: Freedman, Richard
: TITLE OF INVENTION: No. 6440685e1 motor proteins and methods for
: FILE REFERENCE: 1055
: CURRENT APPLICATION NUMBER: US/09/721,689
: CURRENT FILING DATE: 2000-11-24
: NUMBER OF SEQ. ID NOS: 4
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 1375
: TYPE: PR1
: ORGANISM: Human
US-09-721-689-2

Query Match 3.1%; Score 136.5; DB 4; Length 1375;
Best Local Similarity 18.7%; Pred. No. 0.0033;
Matches 154; Conservative 131; Mismatches 299; Indels 239; Gaps 38;

QY 2 DFSWAGAAVNEKSGRTISLSLFQVFAQIFPQWRKGNTECLPYKCSFTGALGENYS 61

```

```

Db 59 DSEFYSA----DTRSPDYVSOEMVFKTLGTDVYKSAFEGYNACVFAV-----GOTSGSKS 109
QY 62 WOIPINHDFILKNNERLOCEVLONKF-----GCISTLVSPVCGNSKSLQVFPKM 113
Db 110 YTMGNNSDGLIP-----RICEGLFSKINETTRMDASFRLEVSLLEYNERVRLLRK 165
QY 114 LTPRIELSV-----WKDDLJTHAVDAVYVNAANEDLLHGGGLALALVKAAGFEIOES 165
Db 166 SSKTFENLREHPKRGYPVEDLSKLYQNYGDV---EELMDAGNINRTTAAAGMNDVSSRS 223
QY 166 KQFARIVKVSAGELAVTGA---GRLPCKQI--IHAVPRMME-WDKQCGT-----GK 212
Db 224 HAIFV-----IKFTQAKFDSMPCTVSKIHLYDLASERDADATGATGVRLEBGN 274
QY 213 LQRAIV-----SILNVLYKNT--HIKTVAPALSSGJFOPLNCT--- 252
Db 275 INKSLVITGAKKQVYFVRDSVLTWLLKDSLGNSKTIMATISPADVNGETILSTLRV 334
QY 253 -----KTIYETIRVSL--QCKPMMSNLKEIHVSN-----EDPTVAFAKASEFI 295
Db 335 ANRAKNIINKPLINEDAVVKLIRELRAIARLKTLLAOGNOLALDSPALSMEELQON 394
QY 296 LGR-SELGOETTPSFNAN--VYNNITLOIYOGHTEMOTADYIVNSVNHDTYGVVASI 352
Db 395 EARVOELTKENTNKNMETONILKEOTLARKEGI---GVVLDSELFHLI----- 440
QY 353 LQAGVEMKSEFLATKAKOFORSOLVLTGKFNLECKYIVLVHSEF-----PKPOI 405
Db 441 ----GID--DDLSTGI-----ILYHKEGQTVGHDASTBEDI 474
QY 406 LKHAKECELECIENOT--TSTISPALGT---GMELTKKET---AAELIFDEVLTAK 455
Db 475 VLHGIDLESEHCIFENIGITVTLPLSGSOSCVNGQIVEATHLNOGAVILLGRTNMPF 534
QY 456 DHVKHQLVFKVIFPTDLEIKAPSEMAKRSKMLSNNSVPOSTREKREN----- 508
Db 535 NHPRE-----AAKREKRKSGLSSFSLSMTDLSKSRNLSAVMLY 575
QY 509 --GLE-ARSPAINLMGFN-----VEENYEAH--AMTORI-----LSIQ 541
Db 576 NPGLEFERQORRELEKLSKRRLIEMEMEKQSKAKALEBMOQOEYTORKEIYQDLIR 635
QY 542 NHHIENNHHILYKKEKHDITISQLOKTSVSTETIISPGRIELEECARADILEVWMI 601
Db 636 KOESILKRRSFHIENKILDLAEKKEFEEERLRE-----OOEIELOKKROE-----E 682
QY 602 DMLCKVOEMARKKRGMLRSLGOWTI-----QOQKTODEKMENTIIFLKCVPPTQELLD 656
Db 683 ETEFLVQOELOLKEPLNNNEKAERFOIPELDOLKEKDEQYAKLELEK-----KRLSE 736
QY 657 OKKOFKCGLOVLYKVEKIDNEVYLAAPORKKKMEKELHROPV 699
Db 737 QEKE-----QVMLVAHLEBOL-----REKQEMTOLLRRGEV 767

RESULT 8
US-09-350-982C-5
Sequence 5, Application US/09350982C
Patent No. 6455290
GENERAL INFORMATION:
APPLICANT: Bertheisen, Jens
APPLICANT: Toma, Salvatore
APPLICANT: Isaac, Antonella
TITLE OF INVENTION: Tankyrase Homolog Protein(THP), Nucleic Acids, and Methods Rel
FILE REFERENCE: PHM-0043
CURRENT APPLICATION NUMBER: US/09/350, 982C
CURRENT FILING DATE: 1999-07-09
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patentin version 3.0
SEQ ID NO 5
LENGTH: 1166
TYPE: PRT

```

```

ORGANISM: Artificial
FEATURE:
NAME/KEY: misc:feature
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc:feature
LOCATION: (1102)..(1102)
OTHER INFORMATION: n is any nucleic acid
NAME/KEY: misc:feature
LOCATION: (2650)..(2650)
OTHER INFORMATION: n is any nucleic acid
US-09-350-982C-5

Query Match
Best Local Similarity 3.0%; Score 134; DB 4; Length 1166;
Matches 43; Conservative 28; Mismatches 74; Indels 24; Gaps 6;

QY 668 VLKVEKIDNEVYLAAPORKKKMEKELHROPVSHRLEFOOVYPCNVYCRVGFQRMSTP 727
Db 994 ILTKIQVCKNKLWERYTHRRKREVESENNHANERMLFPGSP--FVNALIRKGFERNAY- 1050
QY 728 CDPKYGAGIYFTKLNKLNAEKAKRTSAA-----DKLIYFEAEVLGFGFC----- 772
Db 1051 IGMFGAGIYFAENSCKSNQYVYIGGCGCPVHKDRSCYICHRQL--RCRYTLKSF 1107
QY 773 -QGAPLNVPPPLSPGALGDSDVDVNSPSEFVYFSGMOAIPQYLT 820
Db 1108 LQFSAMKVAHSPGHHSVTGRPSV--NGLALAEVYVIRGEQAYPEYLIT 1154

RESULT 9
US-08-973-462-8
Sequence 8, Application US/08973462B
Patent No. 6191270
GENERAL INFORMATION:
APPLICANT: DUBBERSIES, PIERRE
APPLICANT: DRUTHE, PIERRE
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
FILE REFERENCE: 0660-0125-0 PCT
CURRENT APPLICATION NUMBER: US/08/973,462B
CURRENT FILING DATE: 1998-02-06
EARLIER APPLICATION NUMBER: PCT/FR96/00894
EARLIER FILING DATE: 1996-06-12
EARLIER APPLICATION NUMBER: FR 95/07007
EARLIER FILING DATE: 1995-06-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 8
LENGTH: 1786
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

Query Match
Best Local Similarity 3.0%; Score 133; DB 4; Length 1786;
Matches 124; Conservative 128; Mismatches 264; Indels 156; Gaps 28;

QY 81 LCEVLONKFGCISTLVSPVOEGNSKSLQVFRKMLPRLIELSWKDDLTTHAVDAVYVNAAN 140
Db 845 LNEIEVEKENVVTTLENEVEETTAESVTFSSNILE-----EIOENTTINDTIEEKL---- 895
QY 141 EDLLHGGGLALALVKAAGFEIOESKQFARVARKVSAGEIAVTGARGLPCKOIIHAVGPR 200
Db 896 -ELHNNVLSALENT---QSEEEKREYIDVIEYKE--EVAT-----LIETV--- 938
QY 201 WMEWDKQCGTGKQRAIVSILNVYLYKTHIKTVAIIPALSSGIFOPPLNCTKTIYETIR 260
Db 939 --EQAEKKSANTITEIFEENLAEVNSNEN--AENLEKINLETVEFTVLDKVEET--VEISG 994
QY 261 VSLQCKPM-----MSLTKRIHLVSNEDPVAAPFAKASEIILKSELGQETTPSFNAN 312
Db 995 ESLENNEMDKAFSEIFDVKGI---QENLLTGMRFSRTIETSIYQSEKVDLNNVSS 1050

```

QY 313 VVNNLTLQVGHIEWQADVIANSVPHDITVG---PVAKSILQOAGVE-----MKSPF 364
Db 1051 ILDNIE-----NMKEGLNKLNTENISSTEGVQETVTEHVEQNVYVDVVPAMKDOF 1100
QY 365 LATKAKQFORSLVLTGK-----FNL--FCKYIYHVLHNSFPKPOLIKHAMKECLF 415
Db 1101 LG-----ILNEAGGLKEMEFNLEDVFKSESQVTVTEIKDEPVQKEKETVSI 1149
QY 416 -KCIBONITISIFPALGNGMFKETAEILFD---EVLTRAKDHVKHQLTVKVFIFPT 471
Db 1150 IEMENIVDV-----LEEKEDLTKMIDAVEESIEISSKETEESIKKEKEDV 1200
QY 472 DLEIKAFSESEMAKR-SKMLSLNNYSVPOSTREKRENGLEARSPLAINMGFVEMWEA 530
Db 1201 SLVVEVDNDMDSEVKEVLKLN-----MEELMKDAVELINDITSKIE-ETDELNV 1253
QY 531 HA-----WIORIISLONHIIIE--NNHILYLRKREHDITSLQKTTSSVSTIEIIS 578
Db 1254 EADLIKMEKLEKELKALSESEKELIDAKDITLKEVIEEHDTITLDEV----- 1303
QY 579 PGRTELEGARADIEVVMNEDMLCKVQEMARKERGLMRSLGQWTIOQKTKODEMK 638
Db 1304 -----VELKDVEEDKIEKVSDIKLDEEDILKEVKEIKR-----LESEIL 1342
QY 639 ENIFLKCPVPPTQELDQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 689
Db 1343 EDYKELK---TETDILKEKKEIKEDHFEKEFEAEIKDLADLKEVSSLEVEEKKL 1399
QY 690 MEKHLKROPVSH 701
Db 1400 EEVHELKEEVER 1411

RESULT 10
US-08-801-263A-9

Sequence 9, Application US/08801263A
Patent No. 5811407
GENERAL INFORMATION:
APPLICANT: Johnston, Robert E.
APPLICANT: Davis, Nancy L.
APPLICANT: Simpson, Dennis A.
TITLE OF INVENTION: System for the In Vivo Delivery and
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
STREET: 1211 East Morehead Street
CITY: Charlotte
STATE: No. 3811407th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/801,263A
FILING DATE: 19-FEB-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Sibley, Kenneth D.
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5470-147
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2512 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-801-263A-9

Query Match 3.0%; Score 132.5; DB 2; Length 2512;
Best Local Similarity 23.0%; Pred. No. 0.021;
Matches 72; Conservative 45; Mismatches 107; Indels 89; Gaps 13;

QY 26 FCKYVAQIFPCQKRRKQNTCECLPY-----KCSFTGALGENTSMQIPINHDFKILKN----- 76
Db 1227 YDLVFINIGTKYRNHHFQCCEDHAATLEKTLRSALN-----CLNPGCTLVKSGYAD 1279
QY 77 -NEROLCEVLQKFGCISTVSPVOEGNSKSLQVPRKM-----LTPR---ILSYW- 123
Db 1280 RNSSEVYALARKFVRVGAARDQVSSNTMELITRQDLSNSTRQPTFHNHLCVSSVYE 1339
QY 124 -----KDLITHAVDVYVNAANEDILHGGGLALVLKAGCEIODESKOF 168
Db 1340 GTRDVGGAAPSYRTKRENIADQEEAVVNAANPLGRPEGVCRAIYK----- 1386
QY 169 VARYKVGAGELAVYGAQRLP-C--KQIITHAVGPRMMDKQCGTGLQRAVSLNTVYI 225
Db 1387 --RWPTSTTDSATETGTARMVYCLGKQVIAHVGDPFRKHPEALKLQNAVHAYADLV- 1443
QY 226 YKNTH-IKTVAIPALSGIF-----OEPNLICT-----KITIVETRV 261
Db 1444 --NEHNIKSVAILPLSLTGIYVAGDRLEVSINCLTTALDRTDADVTYICLDKMKKERIDA 1501
QY 262 SLOGKPMMSNIKE 274
Db 1502 ALOLKESVTELEKD 1514

RESULT 11
US-09-102-248-9

Sequence 9, Application US/09102248
Patent No. 6008035
GENERAL INFORMATION:
APPLICANT: Johnston, Robert E.
APPLICANT: Davis, Nancy L.
APPLICANT: Simpson, Dennis A.
TITLE OF INVENTION: System for the In Vivo Delivery and
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
STREET: 1211 East Morehead Street
CITY: Charlotte
STATE: No. 6008035th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/102,248
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/801,263
FILING DATE: 19-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sibley, Kenneth D.
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5470-147
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2512 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-102-248-9

Query Match 3.0%; Score 132.5; DB 3; Length 2512;
Best Local Similarity 23.0%; Pred. No. 0.021;
Matches 72; Conservative 45; Mismatches 107; Indels 89; Gaps 13;

26 FQKVAQIPQWRKNGTECLPY-----KCEETGALGENTSWQIPINHNDFILKN----- 76
1227 YDLVFINIGTKRNNHFGQCEDEHAATLKLSSALN-----CLNPGGITLVKSKGYAD 1279
77 -NEROLCEVLQNKFGCISTFLVSPVQEGNSKSLQVERKM-----LTPR-----IELSVN- 123
1280 RNSEDDVTLARKFVNSARPDCVSSNTEMYLIFROLDSRTRQETPHHLNCVISVYE 1339
124 -----KDDLTHAVDAVYVNAANEDLHGGGLALALVKAGFELQESKOF 168
1340 GTRDVGAAPSYRTKRENIADCOEEAVVNAANPLGRGEGVCRAIYK----- 1386
169 VARYGVSAGEIATVAGRLP-C--KQIIHAVGPRMENDKOGCTGKLORAIVSILNYI 225
1387 --RWPSTDSATETGTATMTVCIGKKVTHAVGPDRKHPEALKLLQNAVHAVADLV- 1443
226 YKNTN-IKVVAIPALSSGIF-----OPLNLCT-----KTIYETIV 261
1444 --NEHHIKVAIPILSTGIYAAGKDRLEVSINCLTALDRTDADVTIYCLDKMKKERIDA 1501
262 SLOGKPMNSLKE 274
1502 ALQKESVTELD 1514

RESULT 12
US-09-196-387-2
Sequence 2, Application US/09196387
Patent No. 6277613

GENERAL INFORMATION:

APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
TITLE OF INVENTION: OF USE THEREOF
NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/196.387

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/095,225

FILING DATE: June 10, 1998

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

INFORMATION FOR SEQ. ID NO. 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1327 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-196-387-2

Query Match 3.0%; Score 132; DB 4; Length 1327;
Best Local Similarity 21.0%; Pred. No. 0.0081;
Matches 65; Conservative 48; Mismatches 107; Indels 90; Gaps 15;

558 EHDILSLOKTSSTVITELISPGTELELEGARA--DLIEVVMNEDMLCKVOEEMARKK 615
1041 EH--LNDIETEQITLDVLADMGHELEKEIGAINAGHRKLIKGVRL-----GG 1089
616 ERG-----LMRSLGQWTI-----OQOKODEKKEVITFLKCPVPTQELDQKOF 661
1090 QOQTNEYLFHCVNOQTILLDLAPEDKEVSEEE-----QSTIREHRDG 1135
662 ERG-----LOYLKVEKIDNEVLMMAFORKKKMEKELROPVSHRLFQOVPYFCNVYC 716
1136 GNAGGIFNNYVIRIQKVVNKKLRERFCHROKVESENNHHNRYLFGSP--FINALI 1193
717 RVFORMISTPCDDPKYGAGIYFTKLNKLAERAKKISAA-----DKLIYFEAEVLTG 769
1194 HKGFERNHAY-IGMGAGIYFAENSCKSNQYVYGIGGTCGPTKDRSCYICHRQML-- 1250
770 FFCQ-----GPLENIYPPPLSPGALDGHDSVDVNS---SPETFYIFSG 810
1251 -FCRYTLGKSPLOFSTMKNAH---AP-----GHHSVIGRPSVNCGLAIAEVIYING 1297
811 MOAIPQYLT 820
1298 EOAYPEYLT 1307

RESULT 13
US-08-801-263A-5

Sequence 5, Application US/08801263A
Patent No. 5811407

GENERAL INFORMATION:

APPLICANT: Johnston, Robert E.

APPLICANT: Davis, Nancy L.

APPLICANT: Simpson, Dennis A.

TITLE OF INVENTION: System for the In Vivo Delivery and

TITLE OF INVENTION: Expression of Heterologous Genes in the Bone Marrow

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Bell Seltzer Park & Gibson, P.A.

STREET: 1211 East Morehead Street

CITY: Charlotte

STATE: No. 5811407th Carolina

COUNTRY: USA

ZIP: 28234

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/801,263A

FILING DATE: 19-FEB-1997

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Sibley, Kenneth D.

REGISTRATION NUMBER: 31,665

REFERENCE/DOCKET NUMBER: 5470-147

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-420-2200

TELEFAX: 919-881-3175

INFORMATION FOR SEQ. ID NO. 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 2517 amino acids

TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-801-263A-5

Query Match
 Best Local Similarity 19.5%; Pred No. 0.029; Length 2517;
 Matches 123; Conservative 85; Mismatches 216; Indels 208; Gaps 26;

```

QY 4 SNVAGAAVNEKSGR--ITSLSLFOKVAFOIFPOMKRGNTTECLPYKCGSEFALGENYS 61
DB 1269 TLVVKSYGADRNSEDVYALARKFVRVSA--ARPECVSNTTEMYLIFR-----QDINSRT 1322
QY 62 WQIPINHNDFILKNNEROLCEVLONKFGCISTLVSPVOEGNSKSLQYFRKMLTPRIELS 121
DB 1323 RQFTPRH-----LNCVTSYVEGTRDVGGAAPSYRTKR----- 1355
QY 122 VKKDLTHAVDAVYVNAANEDLLHGGGLALALVYKAGFEIOESKOFVARGKVSAGEIA 181
DB 1356 ---ENIADCOEAVYVNAANPLGRPEGVCRAIYK-----RWFNSTDSAT 1397
QY 182 VTGAGRLP-C--KQIIHAVGPRMMEWDKOGCTGKLORAIVSLNVYIKNTN-IKTVAP 237
DB 1398 ETGTAKLTVCGCKKVIYHAGPDRKHPREAEALKLLQNAVHAAVDLY---NEHNKISVALP 1454
QY 238 ALSSGIF-----QEPNLCTKTIYETIRVSLQCKPMMSNLKEIHLVSNEDPTVAAP-K 289
DB 1455 LLSTGIYAAGKDRLEVSINCLTALDRT-----DADVTTCYCDK 1493
QY 290 AASEFILKSELGQETTPSFNAMYVNNLTLOIVOGHIEMQTADVIVNS---VNPHDITV 345
DB 1494 KWKERT-----DAVLOKESVIELKDEDMEDIDELVWIHPDCLK 1533
QY 346 GPVAKSILOQAGVEMKSEFLATKAKOFORSOLVLTGKFNLFCKYIYVLMHSE----- 399
DB 1534 G-----RKGFSTTKGKLY-----SYFEGTKRFQAAKDMAE 1563
QY 400 ---PPKPO-----ILKHAMKECLEKC-IEONITSI---SPFAGTGMEIKKEIA 442
DB 1564 IKVLEFPNDOESNOLCAVILGETMAIREKCPVDHNPSSPPKTLPCLOMAMTPEP--- 1620
QY 443 AEILFDEVLTFAKHVHQTLVKEVIFPTDLEIYKAFSSSEMAKRSKMLSNY----- 495
DB 1621 ---VHRLRSNNVK-EVT---VCSSTPLPKVIKKNVQKQVCTKVLFNHPHTAFVPA 1669
QY 496 -----SYPOSTREKRENGLEAKSPAINLMGFVNEEYEAHAMIQRLISLQNHII 545
DB 1670 RKYIAPQOPAPPAQAEAPVAAATPPPADNTSLDVTD-----ISLDMEDS 1718
QY 546 IENNHIILYGRKREHDILSLOKTSSVITEII 577
DB 1719 SEGSLFSSFGSDNSTITSMDSWSSGPPSSLEY 1750

```

RESULT 14
 US-09-102-248-5
 Sequence 5, Application US/09102248
 Patent No. 6008035
 GENERAL INFORMATION:
 APPLICANT: Johnston, Robert E.
 APPLICANT: Davis, Nancy L.
 APPLICANT: Simpson, Dennis A.
 TITLE OF INVENTION: System for the in vivo Delivery and
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
 STREET: 1211 East Morehead Street
 CITY: Charlotte
 STATE: No. 6008035th Carolina
 COUNTRY: USA
 ZIP: 28234
 COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/102,248
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/801,263
 FILING DATE: 19-FEB-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Sibley, Kenneth D.
 REGISTRATION NUMBER: 31,665
 REFERENCE/DOCKET NUMBER: 5470-147
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-420-2200
 TELEFAX: 919-881-3175
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2517 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-102-248-5

Query Match
 Best Local Similarity 19.5%; Pred. No. 0.029; Length 2517;
 Matches 123; Conservative 85; Mismatches 216; Indels 208; Gaps 26;

```

QY 4 SNVAGAAVNEKSGR--ITSLSLFOKVAFOIFPOMKRGNTTECLPYKCGSEFALGENYS 61
DB 1269 TLVVKSYGADRNSEDVYALARKFVRVSA--ARPECVSNTTEMYLIFR-----QDINSRT 1322
QY 62 WQIPINHNDFILKNNEROLCEVLONKFGCISTLVSPVOEGNSKSLQYFRKMLTPRIELS 121
DB 1323 RQFTPRH-----LNCVTSYVEGTRDVGGAAPSYRTKR----- 1355
QY 122 VKKDLTHAVDAVYVNAANEDLLHGGGLALALVYKAGFEIOESKOFVARGKVSAGEIA 181
DB 1356 ---ENIADCOEAVYVNAANPLGRPEGVCRAIYK-----RWFNSTDSAT 1397
QY 182 VTGAGRLP-C--KQIIHAVGPRMMEWDKOGCTGKLORAIVSLNVYIKNTN-IKTVAP 237
DB 1398 ETGTAKLTVCGCKKVIYHAGPDRKHPREAEALKLLQNAVHAAVDLY---NEHNKISVALP 1454
QY 238 ALSSGIF-----QEPNLCTKTIYETIRVSLQCKPMMSNLKEIHLVSNEDPTVAAP-K 289
DB 1455 LLSTGIYAAGKDRLEVSINCLTALDRT-----DADVTTCYCDK 1493
QY 290 AASEFILKSELGQETTPSFNAMYVNNLTLOIVOGHIEMQTADVIVNS---VNPHDITV 345
DB 1494 KWKERT-----DAVLOKESVIELKDEDMEDIDELVWIHPDCLK 1533
QY 346 GPVAKSILOQAGVEMKSEFLATKAKOFORSOLVLTGKFNLFCKYIYVLMHSE----- 399
DB 1534 G-----RKGFSTTKGKLY-----SYFEGTKRFQAAKDMAE 1563
QY 400 ---PPKPO-----ILKHAMKECLEKC-IEONITSI---SPFAGTGMEIKKEIA 442
DB 1564 IKVLEFPNDOESNOLCAVILGETMAIREKCPVDHNPSSPPKTLPCLOMAMTPEP--- 1620
QY 443 AEILFDEVLTFAKHVHQTLVKEVIFPTDLEIYKAFSSSEMAKRSKMLSNY----- 495
DB 1621 ---VHRLRSNNVK-EVT---VCSSTPLPKVIKKNVQKQVCTKVLFNHPHTAFVPA 1669
QY 496 -----SYPOSTREKRENGLEAKSPAINLMGFVNEEYEAHAMIQRLISLQNHII 545
DB 1670 RKYIAPQOPAPPAQAEAPVAAATPPPADNTSLDVTD-----ISLDMEDS 1718
QY 546 IENNHIILYGRKREHDILSLOKTSSVITEII 577
DB 1719 SEGSLFSSFGSDNSTITSMDSWSSGPPSSLEY 1750

```


GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 11:44:37 ; Search time 56 Seconds
(without alignments)
1403.392 Million cell updates/sec

Title: US-09-830-762-2
Perfect score: 4447
Sequence: 1 MDFSVMAGAAAYNEKSGRIT.....PMRPFQHPMRGFAAGSPVD 854

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues
Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*

2: /cgn2_6/ptodata/2/pubppaa/PTCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*

7: /cgn2_6/ptodata/2/pubppaa/PTCTUS_PUBCOMB.pep:*

8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*

10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*

12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*

13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*

14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4393	98.8	856	10	US-09-882-529-4
2	4208.5	94.6	819	10	US-09-882-529-2
3	4175.5	93.9	821	10	US-09-882-529-3
4	912	20.5	179	10	US-09-882-529-7
5	841.5	18.9	169	10	US-09-882-529-5
6	751	16.9	145	10	US-09-882-529-9
7	664	14.9	129	10	US-09-882-529-15
8	626	14.1	121	10	US-09-882-529-11
9	626	14.1	121	10	US-09-882-529-13
10	244.5	5.5	170	10	US-09-882-529-6
11	244.5	5.5	170	10	US-09-882-529-8
12	215.5	4.8	132	10	US-09-882-529-10
13	215.5	4.8	132	10	US-09-882-529-16
14	179.5	4.0	116	10	US-09-882-529-12
15	179.5	4.0	116	10	US-09-882-529-14
16	157	3.5	378	10	US-09-925-301-1208
17	156	3.5	204	9	US-09-731-001-5
18	156	3.5	220	9	US-09-731-001-2
19	156	3.5	258	9	US-09-731-001-4

20	156	3.5	598	9	US-09-731-001-3	Sequence 3, Appl1
21	148	3.3	716	10	US-09-815-242-12403	Sequence 12403, A
22	143	3.2	2492	10	US-09-991-258-3	Sequence 3, Appl1
23	140.5	3.2	250	9	US-10-102-806-517	Sequence 517, App
24	137	3.1	1166	9	US-09-972-115A-6	Sequence 6, Appl1
25	134	3.0	802	9	US-09-964-899-41	Sequence 41, Appl1
26	134	3.0	1074	10	US-09-509-196A-2	Sequence 2, Appl1
27	133	3.0	1786	9	US-09-742-096-3	Sequence 3, Appl1
28	132	3.0	1327	10	US-09-972-115A-8	Sequence 8, Appl1
29	132	3.0	1327	10	US-09-841-835-2	Sequence 2, Appl1
30	132	3.0	1945	9	US-09-927-587-2	Sequence 2, Appl1
31	132	3.0	1979	9	US-09-927-587-4	Sequence 4, Appl1
32	130.5	2.9	465	10	US-09-815-242-5335	Sequence 5335, Ap
33	130	2.9	374	10	US-10-091-483-236	Sequence 236, App
34	130	2.9	374	10	US-09-764-846-236	Sequence 236, App
35	124.5	2.8	1019	10	US-09-801-574-76	Sequence 76, Appl
36	124	2.8	2139	9	US-10-023-219-4	Sequence 4, Appl1
37	124	2.8	2139	10	US-09-727-384-6	Sequence 6, Appl1
38	123.5	2.8	822	10	US-09-981-900B-9	Sequence 9, Appl1
39	123.5	2.8	1333	9	US-09-972-115A-2	Sequence 2, Appl1
40	122.5	2.8	1267	9	US-09-972-115A-4	Sequence 4, Appl1
41	122	2.7	846	9	US-09-893-519A-34	Sequence 34, Appl
42	117.5	2.6	617	10	US-09-815-242-13228	Sequence 13228, A
43	117	2.6	2012	9	US-09-808-602-68	Sequence 68, Appl
44	116	2.6	2037	9	US-09-951-402-3	Sequence 3, Appl1
45	116	2.6	2037	10	US-09-951-401-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-09-882-529-4
Sequence 4, Application US/09882529
Patent No. US20020132317A1
GENERAL INFORMATION:
APPLICANT: Peyman, John A
APPLICANT: da Silva, Antonio
APPLICANT: Hockman, Paula
TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
FILE REFERENCE: 15966-771
CURRENT APPLICATION NUMBER: US/09/882, 529
CURRENT FILING DATE: 2001-09-12
PRIOR APPLICATION NUMBER: 60/211, 565
PRIOR FILING DATE: 2000-06-15
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 856
TYPE: PRT
ORGANISM: Homo sapiens
US-09-882-529-4

Query Match 98.8%, Score 4393; DB 10; Length 856;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 849; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY	1	MDFSVMAGAAAYNEKSGRITSLILFKQVFAIFPQWKGNTDEELPKCSTGALGNY 60	DB	1	MDFSVMAGAAAYNEKSGRTSLILFKQVFAIFPQWKGNTDEELPKCSTGALGNY 60
QY	61	SMQIINNDFFILKNNEHOLCEVQNKFGCSTLVSPOEBSKSLQVFRKMLPRIEL 120	DB	61	SMQIINNDFFILKNNEHOLCEVQNKFGCSTLVSPOEBSKSLQVFRKMLPRIEL 120
QY	121	SVWKDILTTTHAVDAVYVNAEDLHGGGLALALVKAAGFEIOESKQFVARYGSAGEI 180	DB	121	SVWKDILTTTHAVDAVYVNAEDLHGGGLALALVKAAGFEIOESKQFVARYGSAGEI 180
QY	181	AVTGAGRLPCQKQIHAVGRWEMDKGCTGLQRAIYSILNYVYKNTHTKIVAPALS 240	DB	181	AVTGAGRLPCQKQIHAVGRWEMDKGCTGLQRAIYSILNYVYKNTHTKIVAPALS 240

```

QY 241 SGIOFPLNLCTKTIVETIRVSLGCKPMAISNLKEIHLVSNEDPTVAARKASEFTLGKSE 300
DB 241 SGIOFPLNLCTKTIVETIRVSLGCKPMAISNLKEIHLVSNEDPTVAARKASEFTLGKSE 300
QY 301 LGOETTPSFNMAVNNLTLOIVOGHIEWOTADVIANSVNPBDITVGPVAKSILQOAGVEM 360
DB 301 LGOETTPSFNMAVNNLTLOIVOGHIEWOTADVIANSVNPBDITVGPVAKSILQOAGVEM 360
QY 361 KSEFLATKAKOFORSOLVLTGKFNLFCKYIYHVLHSEFPKPOLIKHAMKECLEKCIQ 420
DB 361 KSEFLATKAKOFORSOLVLTGKFNLFCKYIYHVLHSEFPKPOLIKHAMKECLEKCIQ 420
QY 421 NITSISFPALGNGMEIKETAEILFDEVLTF-AKDHVKHQLTVKFIYFPDLEIYKAF 479
DB 421 NITSISFPALGNGMEIKETAEILFDEVLTF-AKDHVKHQLTVKFIYFPDLEIYKAF 480
QY 480 SSEMAKRSKMLSLNNYSVQOSTREKRENGLEARSIPAINLNGFNVEEMVEAHAMTORILIS 539
DB 481 SSEMAKRSKMLSLNNYSVQOSTREKRENGLEARSIPAINLNGFNVEEMVEAHAMTORILIS 540
QY 540 LONHHIENNHIILYGRKREHDLSLOLKTSSVSTIEIISPGTELEIGARADILEVYNN 599
DB 541 LONHHIENNHIILYGRKREHDLSLOLKTSSVSTIEIISPGTELEIGARADILEVYNN 600
QY 600 IEDMLCKVOEEMARKKERGLMRSLGOWTIOOQKTODEKKNITFLKCPVPPTOELLDOKK 659
DB 601 IEDMLCKVOEEMARKKERGLMRSLGOWTIOOQKTODEKKNITFLKCPVPPTOELLDOKK 660
QY 660 QFEKCGLOVLKVEKIDNEVLAFAFORKKKMEKLRQPVSHRLEQOVPYQFCNVVCRVGF 719
DB 661 QFEKCGLOVLKVEKIDNEVLAFAFORKKKMEKLRQPVSHRLEQOVPYQFCNVVCRVGF 720
QY 720 FORNYSPPCDPKYAGIYFTKLNKLAERAKKISAADKLIVFEAEVLITGFCQGHPLNTI 779
DB 721 FORNYSPPCDPKYAGIYFTKLNKLAERAKKISAADKLIVFEAEVLITGFCQGHPLNTI 780
QY 780 VPPPLSPGALDGHDSVDNVSSPETF-VIFSGMOAIPQYLMTCQEOYVOSODYSSGPMRPA 838
DB 781 VPPPLSPGALDGHDSVDNVSSPETF-VIFSGMOAIPQYLMTCQEOYVOSODYSSGPMRPA 840
QY 839 FAQHPWRGFASSGSPVD 854
DB 841 FAQHPWRGFASSGSPVD 856

```

RESULT 2

US-09-882-529-2

Sequence 2, Application US/09882529

Patent No. US20020132317A1

GENERAL INFORMATION:

APPLICANT: Peyman, John A

APPLICANT: da Silva, Antonio

APPLICANT: Hockman, Paula

TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC

FILE REFERENCE: 15966-771

CURRENT APPLICATION NUMBER: US/09/882, 529

PRIOR FILING DATE: 2001-09-12

PRIOR APPLICATION NUMBER: 60/211,565

NUMBER OF SEQ ID NOS: 16

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 819

TYPE: PRT

ORGANISM: Homo sapiens

US-09-882-529-2

Query Match 94.6%; Score 4208.5; DB 10; Length 819;

Best Local Similarity 95.4%; Pred. No. 0;

Matches 815; Conservative 3; Mismatches 1; Indels 35; Gaps 1;

```

QY 1 MDESVAGAAVANEKSGRTISLSLLFQKVFQAIFFPMQRKKGNTTECLPYKCEGTALGENY 60
DB 1 MDESVAGAAVANEKSGRTISLSLLFQKVFQAIFFPMQRKKGNTTECLPYKCEGTALGENY 60
QY 61 SMQIDPNNHNDFFILKNNEROLCEVLONKFGCISTLYSPVOEGNSKSLQVFRMLLPRIEL 120
DB 61 SMQIDPNNHNDFFILKNNEROLCEVLONKFGCISTLYSPVOEGNSKSLQVFRMLLPRIEL 120
QY 121 SVMKDGLTTHAVDAVYNNANEDLLHGGGLALALVAGGFEIOEESKQFVARYKGSAGEI 180
DB 121 SVMKDGLTTHAVDAVYNNANEDLLHGGGLALALVAGGFEIOEESKQFVARYKGSAGEI 180
QY 180 SVMKDGLTTHAVDAVYNNANEDLLHGGGLALALVAGGFEIOEESKQFVARYKGSAGEI 145
DB 181 SVMKDGLTTHAVDAVYNNANEDLLHGGGLALALVAGGFEIOEESKQFVARYKGSAGEI 145
QY 145 SVMKDGLTTHAVDAVYNNANEDLLHGGGLALALVAGGFEIOEESKQFVARYKGSAGEI 145
DB 146 SVMKDGLTTHAVDAVYNNANEDLLHGGGLALALVAGGFEIOEESKQFVARYKGSAGEI 145
QY 146 SVMKDGLTTHAVDAVYNNANEDLLHGGGLALALVAGGFEIOEESKQFVARYKGSAGEI 145
DB 146 SVMKDGLTTHAVDAVYNNANEDLLHGGGLALALVAGGFEIOEESKQFVARYKGSAGEI 145
QY 241 SGIOFPLNLCTKTIVETIRVSLGCKPMAISNLKEIHLVSNEDPTVAARKASEFTLGKSE 300
DB 241 SGIOFPLNLCTKTIVETIRVSLGCKPMAISNLKEIHLVSNEDPTVAARKASEFTLGKSE 300
QY 301 LGOETTPSFNMAVNNLTLOIVOGHIEWOTADVIANSVNPBDITVGPVAKSILQOAGVEM 360
DB 301 LGOETTPSFNMAVNNLTLOIVOGHIEWOTADVIANSVNPBDITVGPVAKSILQOAGVEM 360
QY 361 KSEFLATKAKOFORSOLVLTGKFNLFCKYIYHVLHSEFPKPOLIKHAMKECLEKCIQ 420
DB 361 KSEFLATKAKOFORSOLVLTGKFNLFCKYIYHVLHSEFPKPOLIKHAMKECLEKCIQ 420
QY 421 NITSISFPALGNGMEIKETAEILFDEVLTF-AKDHVKHQLTVKFIYFPDLEIYKAF 480
DB 421 NITSISFPALGNGMEIKETAEILFDEVLTF-AKDHVKHQLTVKFIYFPDLEIYKAF 480
QY 480 SSEMAKRSKMLSLNNYSVQOSTREKRENGLEARSIPAINLNGFNVEEMVEAHAMTORILIS 539
DB 481 SSEMAKRSKMLSLNNYSVQOSTREKRENGLEARSIPAINLNGFNVEEMVEAHAMTORILIS 540
QY 540 LONHHIENNHIILYGRKREHDLSLOLKTSSVSTIEIISPGTELEIGARADILEVYNN 599
DB 541 LONHHIENNHIILYGRKREHDLSLOLKTSSVSTIEIISPGTELEIGARADILEVYNN 600
QY 600 IEDMLCKVOEEMARKKERGLMRSLGOWTIOOQKTODEKKNITFLKCPVPPTOELLDOKK 659
DB 601 IEDMLCKVOEEMARKKERGLMRSLGOWTIOOQKTODEKKNITFLKCPVPPTOELLDOKK 660
QY 660 QFEKCGLOVLKVEKIDNEVLAFAFORKKKMEKLRQPVSHRLEQOVPYQFCNVVCRVGF 719
DB 661 QFEKCGLOVLKVEKIDNEVLAFAFORKKKMEKLRQPVSHRLEQOVPYQFCNVVCRVGF 720
QY 720 FORNYSPPCDPKYAGIYFTKLNKLAERAKKISAADKLIVFEAEVLITGFCQGHPLNTI 779
DB 721 FORNYSPPCDPKYAGIYFTKLNKLAERAKKISAADKLIVFEAEVLITGFCQGHPLNTI 780
QY 780 VPPPLSPGALDGHDSVDNVSSPETF-VIFSGMOAIPQYLMTCQEOYVOSODYSSGPMRPA 838
DB 781 VPPPLSPGALDGHDSVDNVSSPETF-VIFSGMOAIPQYLMTCQEOYVOSODYSSGPMRPA 840
QY 839 FAQHPWRGFASSGSPVD 854
DB 841 FAQHPWRGFASSGSPVD 856

```

RESULT 3

US-09-882-529-3

Sequence 3, Application US/09882529

Patent No. US20020132317A1

GENERAL INFORMATION:

APPLICANT: Peyman, John A

APPLICANT: da Silva, Antonio

APPLICANT: Hockman, Paula

TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC

FILE REFERENCE: 15966-771

CURRENT APPLICATION NUMBER: US/09/882, 529

; CURRENT FILING DATE: 2001-09-12
 ; PRIOR APPLICATION NUMBER: 60/211,565
 ; PRIOR FILING DATE: 2000-06-15
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 821
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-882-529-3

Query Match 93.9%; Score 4175.5; DB 10; Length 821;
 Best Local Similarity 95.1%; Pred. No. 0;
 Matches 814; Conservative 0; Mismatches 5; Indels 37; Gaps 3;

QY 1 MPFSWAGAAAYNEKSGRTISLSLFQKFAQIFPQMRKGNTECLPYCSGALGENY 60
 DB 1 MPFSWAGAAAYNEK-----SETGALGENY 25
 QY 61 SMOIPINNDPFLIKKNEBOLCEVLONKPGCISTLWSPVOEGNSKSLQVFRKKLTFRIL 120
 DB 26 SMOIPINNDPFLIKKNEBOLCEVLONKPGCISTLWSPVOEGNSKSLQVFRKKLTFRIL 85
 QY 121 SYWKDDLTTHAVDAVVAANEDLLHGGGLALALVKGAFEIQEESKQFVARYGKSAGEI 180
 DB 86 SYWKDDLTTHAVDAVVAANEDLLHGGGLALALVKGAFEIQEESKQFVARYGKSAGEI 145
 QY 181 ATYGAGRLCKCOIHAHVGRMMEMDKOGCTGKLOAIVSLNVIYKNTHTVAIPAL 240
 DB 146 ATYGAGRLCKCOIHAHVGRMMEMDKOGCTGKLOAIVSLNVIYKNTHTVAIPAL 205
 QY 241 SGIFQPLNLCRTIVETIRVSLQGRPMNSNKEIHLVSNEDPTVAAFRAASEFLIGKSE 300
 DB 206 SGIFQPLNLCRTIVETIRVSLQGRPMNSNKEIHLVSNEDPTVAAFRAASEFLIGKSE 265
 QY 301 LQOETTPSNANVNNLTQIYQGHLEMTADYVNSVPHDITVGPVAKSIIQAGVEM 360
 DB 266 LQOETTPSNANVNNLTQIYQGHLEMTADYVNSVPHDITVGPVAKSIIQAGVEM 325
 QY 361 KSEFLATKAKOFORSOLVLTGKFNLCFYIYVHLWHSFEPKPOLKHNMECELECIQ 420
 DB 326 KSEFLATKAKOFORSOLVLTGKFNLCFYIYVHLWHSFEPKPOLKHNMECELECIQ 385
 QY 421 NITSISFPALGTGNMEIKETAELFDEVLTFF-ANDVYKHOLTVKVFJPTDLEITYKAF 479
 DB 386 NITSISFPALGTGNMEIKETAELFDEVLTFF-ANDVYKHOLTVKVFJPTDLEITYKAF 445
 QY 480 SSEMARSKMLSLNNYSVQSTREKRENGLEKSPAINLMGFNVSENEEAHAMIORIIS 539
 DB 446 SSEMARSKMLSLNNYSVQSTREKRENGLEKSPAINLMGFNVSENEEAHAMIORIIS 505
 QY 540 LONHHIENNHLIYLGRRKHDILISOKTSSVSITELISPGFTELEIEGARADLIEVNN 599
 DB 506 LONHHIENNHLIYLGRRKHDILISOKTSSVSITELISPGFTELEIEGARADLIEVNN 565
 QY 600 IEDMLCKVOEBARKKREKLMSLQGTIOQKTODEMKNILFLKCPVPPTOELLDOKK 659
 DB 566 IEDMLCKVOEBARKKREKLMSLQGTIOQKTODEMKNILFLKCPVPPTOELLDOKK 625
 QY 660 OFEKGLQVLYKVEKIDNEVLMMAFQRRKKMEKHLRQVSHRLPQOYVQFCNVYCRIG 719
 DB 626 OFEKGLQVLYKVEKIDNEVLMMAFQRRKKMEKHLRQVSHRLPQOYVQFCNVYCRIG 685
 QY 720 FORMYSTPCDPKYGAGIYFTKLNKLAERAKKISADKLIVYEAVLITGFCOGHPLNT 779
 DB 686 FORMYSTPCDPKYGAGIYFTKLNKLAERAKKISADKLIVYEAVLITGFCOGHPLNT 745
 QY 780 VPPPLSPGALIDGHDVAVNVSSPEF-VIFSGMOAIPQYLTWCTQYVOSQDYSSGPMRP 838
 DB 746 VPPPLSPGALIDGHDVAVNVSSPEF-VIFSGMOAIPQYLTWCTQYVOSQDYSSGPMRP 805
 QY 839 FAQHPWRGFGASGSPVD 854
 |||||||||||||||

DB 806 FAQHPWRGFGASGSPVD 821

RESULT 4
 US-09-882-529-7
 ; Sequence 7, Application US/09882529
 ; Patent No. US20020132317A1
 ; GENERAL INFORMATION:

; APPLICANT: Peyman, John A
 ; APPLICANT: da Silva, Antonio
 ; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
 ; FILE REFERENCE: 15966-771
 ; CURRENT APPLICATION NUMBER: US/09/882,529
 ; CURRENT FILING DATE: 2001-09-12
 ; PRIOR APPLICATION NUMBER: 60/211,565
 ; PRIOR FILING DATE: 2000-06-15
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 7
 ; LENGTH: 179
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-882-529-7

Query Match 20.5%; Score 912; DB 10; Length 179;
 Best Local Similarity 99.4%; Pred. No. 2,7e-64;
 Matches 178; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 114 LTPRIELSYWKDDLTTHAVDAVVAANEDLLHGGGLALALVKGAFEIQEESKQFVARYG 173
 DB 1 LTPRIELSYWKDDLTTHAVDAVVAANEDLLHGGGLALALVKGAFEIQEESKQFVARYG 60
 QY 174 KVSAGEIATYGAGRLCKCOIHAHVGRMMEMDKOGCTGKLOAIVSLNVIYKNTHT 233
 DB 61 KVSAGEIATYGAGRLCKCOIHAHVGRMMEMDKOGCTGKLOAIVSLNVIYKNTHT 120
 QY 234 VAIPALSSGIFQPLNLCRTIVETIRVSLQGRPMNSNKEIHLVSNEDPTVAAFRAAS 292
 DB 121 VAIPALSSGIFQPLNLCRTIVETIRVSLQGRPMNSNKEIHLVSNEDPTVAAFRAAS 179

RESULT 5
 US-09-882-529-5
 ; Sequence 5, Application US/09882529
 ; Patent No. US20020132317A1
 ; GENERAL INFORMATION:

; APPLICANT: Peyman, John A
 ; APPLICANT: da Silva, Antonio
 ; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
 ; FILE REFERENCE: 15966-771
 ; CURRENT APPLICATION NUMBER: US/09/882,529
 ; CURRENT FILING DATE: 2001-09-12
 ; PRIOR APPLICATION NUMBER: 60/211,565
 ; PRIOR FILING DATE: 2000-06-15
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 169
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-882-529-5

Query Match 18.9%; Score 841.5; DB 10; Length 169;
 Best Local Similarity 97.6%; Pred. No. 9,1e-59;
 Matches 165; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 313 VVNNLTLOIVQGHIEMTADYVNSVPHDITVGPVAKSIIQAGVEMKSEFLATKAKOF 372
 DB 1 VVNNLTLOIVQGHIEMTADYVNSVPHDITVGPVAKSIIQAGVEMKSEFLATKAKOF 60
 |||||||||||||||

Qy 373 ORSOLVLTGKFNLFCKYIYHVMHSEFPKPOLKHKAMECKLEKCIQNITISIPALGT 432
|||||
Db 61 QRSOLVLTGKFNLFCKYIYHVMHSEFPKPOLKHKAMECKLEKCIQNITISIPALGT 120
Qy 433 GMMETKKEEAAILDFDEVLT-F-AKDHVKHQLTVKRVIPFTDLEIKAS 480
|||||
Db 121 GMMETKKEEAAILDFDEVLTAKDHVKHQLTVKRVIPFTDLEIKAS 169

RESULT 6
US-09-882-529-9
: Sequence 9, Application US/09882529
: Patent No. US20020132317A1
: GENERAL INFORMATION:
: APPLICANT: Peyman, John A
: APPLICANT: da Silva, Antonio
: APPLICANT: Hockman, Paula
: TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
: FILE REFERENCE: 15966-771
: CURRENT APPLICATION NUMBER: US/09/882,529
: CURRENT FILING DATE: 2001-09-12
: PRIOR APPLICATION NUMBER: 60/211,565
: PRIOR FILING DATE: 2000-06-15
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 9
: LENGTH: 145
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-882-529-9

Query Match 16.9%; Score 751; DB 10; Length 145;
Best Local Similarity 100.0%; Pred. No. 1e-51;
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 RKMLTPRIELSWKDDLTTHAVDAVYNAANEDLHGGGLALALVYKAGFEIOESKQFYA 170
|||||
Db 1 RKMLTPRIELSWKDDLTTHAVDAVYNAANEDLHGGGLALALVYKAGFEIOESKQFYA 60
Qy 171 RYKGSAGEIAVYAGRIKPKQIIHAYGPRWMEKOCCTGKLQRAIVSILNVYTKNTH 230
|||||
Db 61 RYKGSAGEIAVYAGRIKPKQIIHAYGPRWMEKOCCTGKLQRAIVSILNVYTKNTH 120
Qy 231 IKTYAIPALSSGIFQFPLNLCTKTI 255
|||||
Db 121 IKTYAIPALSSGIFQFPLNLCTKTI 145

RESULT 7
US-09-882-529-15
: Sequence 15, Application US/09882529
: Patent No. US20020132317A1
: GENERAL INFORMATION:
: APPLICANT: Peyman, John A
: APPLICANT: da Silva, Antonio
: APPLICANT: Hockman, Paula
: TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
: FILE REFERENCE: 15966-771
: CURRENT APPLICATION NUMBER: US/09/882,529
: CURRENT FILING DATE: 2001-09-12
: PRIOR APPLICATION NUMBER: 60/211,565
: PRIOR FILING DATE: 2000-06-15
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 15
: LENGTH: 129
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-882-529-15

Query Match 14.9%; Score 664; DB 10; Length 129;
Best Local Similarity 99.2%; Pred. No. 6.1e-45;
Matches 128; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 318 TLQIVQGHLEMOGTADVYNSVNPHTITGPAVAKSILLOAGVPMKSEFLATRAKOFORSOL 377
|||||
Db 1 TLQIVQGHLEMOGTADVYNSVNPHTITGPAVAKSILLOAGVPMKSEFLATRAKOFORSOL 60
Qy 378 VLVTKGFNLFCKYIYHVMHSEFPKPOLKHKAMECKLEKCIQNITISIPALGTGMMEI 437
|||||
Db 61 VLVTKGFNLFCKYIYHVMHSEFPKPOLKHKAMECKLEKCIQNITISIPALGTGMMEI 120
Qy 438 KKEETAAIL 446
|||||
Db 121 KKEETAAIL 129

RESULT 8
US-09-882-529-11
: Sequence 11, Application US/09882529
: Patent No. US20020132317A1
: GENERAL INFORMATION:
: APPLICANT: Peyman, John A
: APPLICANT: da Silva, Antonio
: APPLICANT: Hockman, Paula
: TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
: FILE REFERENCE: 15966-771
: CURRENT APPLICATION NUMBER: US/09/882,529
: CURRENT FILING DATE: 2001-09-12
: PRIOR APPLICATION NUMBER: 60/211,565
: PRIOR FILING DATE: 2000-06-15
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 11
: LENGTH: 121
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-882-529-11

Query Match 14.1%; Score 626; DB 10; Length 121;
Best Local Similarity 100.0%; Pred. No. 5.5e-42;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 136 VNAANEDLHGGGLALALVYKAGFEIOESKQFYARYKVSAGETAVTGAGRLPCKQIIT 195
|||||
Db 1 VNAANEDLHGGGLALALVYKAGFEIOESKQFYARYKVSAGETAVTGAGRLPCKQIIT 60
Qy 196 AVGPRWMEKOCCTGKLQRAIVSILNVYTKNTHIKTYAIPALSSGIFQFPLNLCTKTI 255
|||||
Db 61 AVGPRWMEKOCCTGKLQRAIVSILNVYTKNTHIKTYAIPALSSGIFQFPLNLCTKTI 120
Qy 256 V 256
|
Db 121 V 121

RESULT 9
US-09-882-529-13
: Sequence 13, Application US/09882529
: Patent No. US20020132317A1
: GENERAL INFORMATION:
: APPLICANT: Peyman, John A
: APPLICANT: da Silva, Antonio
: APPLICANT: Hockman, Paula
: TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
: FILE REFERENCE: 15966-771
: CURRENT APPLICATION NUMBER: US/09/882,529
: CURRENT FILING DATE: 2001-09-12
: PRIOR APPLICATION NUMBER: 60/211,565
: PRIOR FILING DATE: 2000-06-15
: NUMBER OF SEQ ID NOS: 16

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-882-529-13
```

```
Query Match
Best Local Similarity 14.1%; Score 626; DB 10; Length 121;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 136 VNAANEDLHGGLALALVAKAGFEIOESKQFVARYGVSAAGELAVTGAIRLPCKQIIT 195
DB 1 VNAANEDLHGGLALALVAKAGFEIOESKQFVARYGVSAAGELAVTGAIRLPCKQIIT 60
OY 196 AVGPFRMMDKOGCTGKLORAIVSILNVIYKNTHTKTAIPALSSGIFOPPLNCTKT 255
DB 61 AVGPFRMMDKOGCTGKLORAIVSILNVIYKNTHTKTAIPALSSGIFOPPLNCTKT 120
OY 256 V 256
DB 121 V 121
```

```
RESULT 10
US-09-882-529-6
; Sequence 6, Application US/09882529
; Patent No. US20020132317A1
; GENERAL INFORMATION:
; APPLICANT: Peyman, John A
; APPLICANT: da Silva, Antonio
; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
; FILE REFERENCE: 15966-771
; CURRENT APPLICATION NUMBER: US/09/882,529
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 60/211,565
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-882-529-6
```

```
Query Match
Best Local Similarity 5.5%; Score 244.5; DB 10; Length 170;
Matches 64; Conservative 31; Mismatches 70; Indels 11; Gaps 4;
```

```
OY 115 TPRIELSVKDDLTTHAVDAVVAANANEDLHGGLALALVAKAGF-EIOESKQFVARYG 173
DB 2 TGSMTKIKVKGDTITKLPAADAVNAANSDLTMGGVAGAIARAAGEPELEEE---LKG 57
OY 174 KVSAGEIAVTGAGRPLCKQIITHAVGPRMMDKOGCTGKLORAIVSILNVIYKNTHTKT 233
DB 58 GVPTEGAVVTPGGNLPKAKYVIAHVGPRMMDKOGCTGKLORAIVSILNVIYKNTHTKT 115
OY 234 VAIPALSSGIFOPPLNCTKTIVETIRVSLQCKPMMSNKEIHLVSNEDPTVAAPK 289
DB 116 VAFPAISTGTIGFPPKDRARIIIEAIRFLTSHA---YKEVVLVLDEEMREAYE 167
```

```
RESULT 11
US-09-882-529-8
; Sequence 8, Application US/09882529
; Patent No. US20020132317A1
; GENERAL INFORMATION:
; APPLICANT: Peyman, John A
; APPLICANT: da Silva, Antonio
; APPLICANT: Hockman, Paula
; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
```

```
; TITLE OF INVENTION: ACIDS ENCODING SAME
; FILE REFERENCE: 15966-771
; CURRENT APPLICATION NUMBER: US/09/882,529
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 60/211,565
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-882-529-8
```

```
Query Match
Best Local Similarity 5.5%; Score 244.5; DB 10; Length 170;
Matches 64; Conservative 31; Mismatches 70; Indels 11; Gaps 4;
```

```
OY 115 TPRIELSVKDDLTTHAVDAVVAANANEDLHGGLALALVAKAGF-EIOESKQFVARYG 173
DB 2 TGSMTKIKVKGDTITKLPAADAVNAANSDLTMGGVAGAIARAAGEPELEEE---LKG 57
OY 174 KVSAGEIAVTGAGRPLCKQIITHAVGPRMMDKOGCTGKLORAIVSILNVIYKNTHTKT 233
DB 58 GVPTEGAVVTPGGNLPKAKYVIAHVGPRMMDKOGCTGKLORAIVSILNVIYKNTHTKT 115
OY 234 VAIPALSSGIFOPPLNCTKTIVETIRVSLQCKPMMSNKEIHLVSNEDPTVAAPK 289
DB 116 VAFPAISTGTIGFPPKDRARIIIEAIRFLTSHA---YKEVVLVLDEEMREAYE 167
```

```
RESULT 12
US-09-882-529-10
; Sequence 10, Application US/09882529
; Patent No. US20020132317A1
; GENERAL INFORMATION:
; APPLICANT: Peyman, John A
; APPLICANT: da Silva, Antonio
; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
; FILE REFERENCE: 15966-771
; CURRENT APPLICATION NUMBER: US/09/882,529
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 60/211,565
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-882-529-10
```

```
Query Match
Best Local Similarity 4.8%; Score 215.5; DB 10; Length 132;
Matches 53; Conservative 18; Mismatches 52; Indels 5; Gaps 3;
```

```
OY 120 LSWMKDDLTTHAVDAVVAANANEDLHGGLALALVAKAGFEIOESKQFVARYGVSA 179
DB 2 LKVKAGDITKPRADAVNAANSDDAHGGVAGAIARAAGE--ESKEEFKRLAGECPVGT 59
OY 180 IAVTGAIRLPCKQIITHAVGPRMMDKOGCTGKLORAIVSILNVIYKNTHTKTAIPAL 239
DB 60 AVTEGCGNLPKAKYVIAHVGPRMMDKOGCTGKLORAIVSILNVIYKNTHTKTAIPAL 116
OY 240 SSGIFOP 247
DB 117 GTGTYGVP 124
```

```
RESULT 13
US-09-882-529-16
```

```

Sequence 16: Application US/09862529
Patent No. US20020132317A1
GENERAL INFORMATION:
APPLICANT: Peyman, John A
APPLICANT: da Silva, Antonio
APPLICANT: Hockman, Paula
TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
FILE OF INVENTION: ACIDS ENCODING SAME
FILE REFERENCE: 15966-771
CURRENT APPLICATION NUMBER: US/09/882,529
CURRENT FILING DATE: 2001-09-12
PRIOR APPLICATION NUMBER: 60/211,565
PRIOR FILING DATE: 2000-06-15
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 16
LENGTH: 132
TYPE: PRT
ORGANISM: Homo sapiens
US-09-862-529-16

```

```

Query Match          4.8%; Score 215.5; DB 10; Length 132;
Best Local Similarity 41.4%; Pred. No. 1.6e-09;
Matches 53; Conservative 18; Mismatches 52; Indels 5; Gaps 3.

QY 120 LSWKDDLTTHAVDVAVNAANEDLLHGGGLALALVAKAGFFETQESKQFVARYGRVSGE 179
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 LKVVAGDITKPRADIVVNAANSDAGHGSGVACALARAAGWE--EKEEERKLAGECPVGT 59

QY 180 IAVTGAAGLPCCKQIILHVGPRMMEKDKGCGTGLORALVSIINVIYKNTHTKYVAIPAL 239
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60 AVVEGGMLPKPKYVIVHAGVPEASGSYKEGYE-LLENVAFACRIAI--ELGTSKVAIPLI 116

QY 240 SSGIQFP 247

Db 117 GTGIVGP 124

```

RESULT 14
 US-09-882-529-12
 Sequence 12, Application US/09882529
 Patent No. US20020132317A1
 GENERAL INFORMATION:
 APPLICANT: Peyman, John A
 APPLICANT: da Silva, Antonio
 APPLICANT: Hockman, Paula
 TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
 FILE REFERENCE: 15966-771
 CURRENT APPLICATION NUMBER: US/09/882,529
 CURRENT FILING DATE: 2001-09-12
 PRIOR APPLICATION NUMBER: 60/211,565
 PRIOR FILING DATE: 2000-06-15
 NUMBER OF SEQ ID NOS: 16
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 12
 LENGTH: 116
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-882-529-12

[illegible]

QY	256 V	256
Db	116 L	116

```

RESULT 15
US-09-882-529-14
: Sequence 14, Application US/09882529
: Patent No. US20020132317A1
: GENERAL INFORMATION:
: APPLICANT: Peyman, John A
: APPLICANT: da Silva, Antonio
: APPLICANT: Hockman, Paula
: TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NOCLETID
: TITLE OF INVENTION: ACIDS ENCODING SAME
: FILE REFERENCE: 15966-771
: CURRENT APPLICATION NUMBER: US/09/882,529
: CURRENT FILING DATE: 2001-09-12
: PRIOR APPLICATION NUMBER: 60/211,565
: PRIOR FILING DATE: 2000-06-15
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: patentln Ver. 2.1
: SEQ ID NO 14
: LENGTH: 116
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-882-529-14

```

	Query Match	4.0%;	Score 179.5;	DB 10;	Length 116;
	Best Local Similarity	37.2%;	Pred. No. 8.8e+07;		
	Matches	45;	Conservative	23;	Mismatches 46; Indels 5; Gaps 2.
OY	136 VNAANEDLLHGGLALALVKAGFEIOESKOFVARYGVNSAGELAVTGARLPCKQIIT	195			
	: : : : : : : : : : :				
Dd	1 VNAANSRLHGGVGAGALARAGKCEAMPEA---FKKAPRCVPYGEAVLTGGGLPAKYVIH	57			
	: : : : : : : : : : :				
OY	136 AVGPPMMENDKCGCTGKLORATIVSLINLYIKNTMIKTVAIPALSSGTFQFPPLNICTKTI	255			
	: : : : : : : : : : :				
Dd	58 AVGPNFSKGCEEEGDELLEKAYRAILR--LADENGCIKSVAFFLLSTGIYGCPKDRAAQSL	115			
	: : : : : : : : : : :				
OY	256 V 256				
Dd	116 L 116				

Search completed: May 13, 2003, 11:54:36
Job time : 59 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 11:39:37 / Search time 49 Seconds
(without alignments)
1675.486 Million cell updates/sec

Title: US-09-830-762-2
Perfect score: 4447
Sequence: 1 MDESMVAGAAAYNEKSGRIT.....PMRPFQHPWRGFGSGSPVD 854
Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_73: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	291.5	6.6	599	2	B72368 conserved hypothet
2	288	6.5	177	2	A99469 conserved hypothet
3	235	5.3	176	2	AF1419 hypothetical prote
4	232	5.2	176	2	AG1794 hypothetical prote
5	220	4.9	173	2	I39569 hypothetical prote
6	211	4.7	173	2	E83182 conserved hypothet
7	204.5	4.6	186	2	AD0636 conserved hypothet
8	203	4.6	190	2	E71027 hypothetical prote
9	194	4.4	169	2	T35937 hypothetical prote
10	189	4.3	359	2	T12540 hypothetical prote
11	187.5	4.2	343	2	H70537 probable lppd prot
12	186.5	4.2	177	2	G90806 probable polypote
13	186.5	4.2	177	2	C85866 probable polypote
14	186.5	4.2	177	2	B64847 probable polypote
15	185	4.2	186	2	H75106 hypothetical prote
16	183.5	4.1	657	2	JC7785 TCDP-inducible pol
17	183	4.1	199	2	H69439 conserved hypothet
18	165.5	3.7	368	2	I80811 histone H2A.1 - ra
19	160	3.6	203	2	T18653 conserved hypothet
20	159.5	3.6	165	2	E70385 conserved hypothet
21	156	3.5	598	2	T46327 hypothetical prote
22	156	3.5	4085	2	S28600 hypothetical prote
23	155	3.5	1169	2	A64505 p115 homolog - Met
24	151	3.4	194	2	D72545 hypothetical prote
25	150.5	3.4	191	2	E84831 conserved hypothet
26	148	3.3	170	2	B75291 conserved hypothet
27	147	3.3	716	2	E89998 conserved hypothet
28	147	3.3	2492	1	C44213 nonstructural poly
29	146	3.3	266	2	G89797 conserved hypothet

30	146	3.3	1552	2	G86344 T22111.2 protein -
31	145	3.3	728	2	A81385 probable ATP/GTP
32	139	3.1	2492	1	MNMVMD nonstructural poly
33	138	3.1	2492	1	A44213 nonstructural poly
34	137	3.1	4488	1	KRIHM2 genome polypote
35	136.5	3.1	689	2	T40364 hypothetical prote
36	136	3.1	2104	2	T38774 myosin-3 heavy cha
37	135	3.0	1899	2	T32732 PAM C-terminal int
38	135	3.0	1919	2	T42098 synaptonemal compl
39	133	3.0	845	2	I48176 RESA-H3 antigen PR
40	133	3.0	1558	2	B71603 nucleotide-binding
41	132.5	3.0	1620	2	S61535 nonstructural poly
42	132.5	3.0	2512	1	MNMVS myosin heavy chain
43	132	3.0	1972	1	A41604 chromosome segrega
44	131.5	3.0	1233	2	T30534 centromere protein
45	131.5	3.0	2663	1	S28261

ALIGNMENTS

RESULT 1

B72368 conserved hypothetical protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: B72368

R:Neelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,

C.M.

Nature 399, 323-329, 1999

A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: B72368

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-599 <ARN>

A:Cross-references: GB:AE001727; GB:AE000512; NID:94981015; PIDN:AD035593.1; PID:9498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0508

Query Match 6.6%, Score 291.5; DB 2; Length 599;
Best Local Similarity 31.9%; Pred. No. 1.8e-10;
Matches 80; Conservative 47; Mismatches 91; Indels 33; Gaps 7;

QY	57	GENYEMQIPININDE-----KILKNN-----EROLCEYLQKKFGCISTLVSPV 99
DB	359	GGGYLY--PHDGGFVKTNLYPEKLNKNEVIFQPKRVGEFEELFERLKIM-----PE 408
QY	100	QEGNSKSLQVPRKMLTPRIETLSVWKDDLTTHAVDAVNAANEDLHGGGLALATVAKAGF 159
DB	409	KYGGSMAMEVRKELEKKKIRIVKGDITREVDIVANANENYKLGSGVACAIYRAGGS 468
QY	160	ETQESKQFVARVYKVSAGEIAVAGARLPCKOIIHAVGPRMWDKQCTGKIORAIVS 219
DB	469	VIOESDRIVDRGRVPTGEAVVTSAGKIKAKYVHTGVPW----RGSSHEDELLYKA 524
QY	220	ILNYIYKN-THIKTVAIPALSSGTFQFPPLNCTTIYETTRVSLQGMNSLNKEIHLV 278
DB	525	VYNALRAHEMLKSTISPAISTGIFGPKERAVGIFSKAIDFDQHP-DITTEIRNIC 583
QY	279	SNEDPTVAAFK 289
DB	584	NIDETTKIFE 594

RESULT 2

A99469 conserved hypothetical protein [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001

C:Accession: A99469

Query March 4.3%, Score 189; DB 2; Length 359;
Best Local Similarity 22.9%; Pred. No. 0.00022;
Matches 80; Conservative 54; Mismatches 89; Indels 126; Gaps 16;

```
OY      536 RILSLQNHIIENN-----HILYLGRKENDLSQLQKTSSVSITE 575  
       | :|::| :| :| :| :| :| :| :| :| :| :| :| :|  
Db      82 REMAMNNHYILLNHSFPRREIKRRLPLFRSCFILPDLQTGLGVPPAPPLEATSS---SQ 138  
OY      576 IISPGRTLELEGARA-----DILEVANNIEDMLCKVGEEMARKKEBGLM 620  
       ||||| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db     139 IICP-----DGVTASANFPETWYMHPSDOFIDVPVASBD-----KSRYITLY 180  
OY      621 RSLGOWTIOQQODTEMEKENIIFLCKCPVRPIQELLQKKCFEGCLGYLVAKVEKDNEYLM 680  
       ||||| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db     181 MLF-----HKTYPEEFKYRI-----LTLLRVQ--NQFLM 206  
OY      681 AAFORKKMKMEEKLHRDP--VSHR-LFOQVPYGFNCNVACRGFORMYSTPCDPK----- 731  
       ::||| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db     207 EKYRKKEYMMRKMGRRIRIINERHLFGHTSODYVDGIKKNF-----DPRVCGRHA 258  
OY      732 --YGAGIFYFTNLK---MLAEKAKKISADRLIVFEAEVLTFGCFCQPLNIIVPPKSP 786  
       :||| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db     259 TMFGGGSYFAKKASYSNHFSKSSK----GVHEHFLAKVLTLGRYTMSHGCMRRRPYPNP 313  
OY      787 GAI--DGHDVVADVNSPETVFVIFSGMAIPOLYMTCTOEYVOOSDYSS 833  
       ||| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db     314 GSVMSDLVDSCVDNFEPQIVFIENDDQSYPYEV-----IQYEVENN 355  
RESULT 11  
#70517  
probable lppd protein - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999  
C:Accession: #70517  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,  
S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
!Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
```

Query Match	4.28	Score 187.5	DB 2	Length 343
Best Local Similarity	29.08	Pred. No. 0.00026		
Matches 56	Conservative 32	Mismatches 74	Indels 31	Gaps 5

Db 334 -----IQAGEDT 340

G90806

```
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
```

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

A;Reference number: A99629; MUID:21156231; PMID:11258796

A: Molecule type: DNA

A;Experimental source: strain O157:H7, substrain RIMD 0509952

C;Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Best Local Similarity 32.28; Pred. No. 0.00011;

QY 117 RIELSVKDDLTTHAVDAVNANEDDLHGGLALALVKAGFEIQEESKQFVARYKVS 176

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 104

Db 62 TGHAVITLACDLPAKAVVHTVGPVW-RGGEQNEEDQLQDAYLNSLRVAANS--YTSVAF 118

[illegible]

probable polyprotein 21679 [imported] - Escherichia coli (strain O157:H7, substrain E

C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: C85666
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Hiller, U.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Diallanata, E.; Polmanousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85666
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <STO>
A:Cross-references: GB:AE005174; NID:912514577; PIDN:AAG55791.1; GSPDB:GN00145; UWGP:Z16
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 21679
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 4.2%; Score 186.5; DB 2; Length 177;
Best Local Similarity 32.2%; Pred. No. 0.00011;
Matches 46; Conservative 29; Mismatches 65; Indels 3; Gaps 2;

OY 117 RLEISVWMDLTTTHAVDVANANEDLLHGGGLALALYKAGGFELQESKQFVARYKVS 176
DB 2 KTRHVVGDITKLAVDIYVNNANPSLMGGGVGDAIHRAPALDACLKVRQOGDCP 61
OY 177 AGEIAVTGAGRLPCQIIHAGVPRMMEMDKQCTGKLORAIVSIINVIYKNTHTIKTVAI 236
DB 62 TGHAVITLADGLPAKAVVHTVGPVW-RGGEONEDDLDQAYINSLRLVAANS--YTSVAF 118
OY 237 PALSGIFQFPLNCTKTIVETI 259
DB 119 PAISTGVYGYPRAAAEIAVKTV 141

RESULT 14

B64847
Probable polypeptide b1045 [similarity] - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C:Accession: B64847

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: B64847

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-177 <BLAT>

A:Cross-references: GB:AE000206; GB:U00096; NID:91787282; PIDN:AAC74129.1; PID:91787283;

A:Experimental source: strain K-12, substrain MG1655

C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 4.2%; Score 186.5; DB 2; Length 177;
Best Local Similarity 32.2%; Pred. No. 0.00011;
Matches 46; Conservative 29; Mismatches 65; Indels 3; Gaps 2;

OY 117 RLEISVWMDLTTTHAVDVANANEDLLHGGGLALALYKAGGFELQESKQFVARYKVS 176
DB 2 KTRHVVGDITKLAVDIYVNNANPSLMGGGVGDAIHRAPALDACLKVRQOGDCP 61
OY 177 AGEIAVTGAGRLPCQIIHAGVPRMMEMDKQCTGKLORAIVSIINVIYKNTHTIKTVAI 236
DB 62 TGHAVITLADGLPAKAVVHTVGPVW-RGGEONEDDLDQAYINSLRLVAANS--YTSVAF 118
OY 237 PALSGIFQFPLNCTKTIVETI 259
DB 119 PAISTGVYGYPRAAAEIAVKTV 141

RESULT 15

H75106

hypothetical protein PAB0445 - Pyrococcus abyssi (strain Orsay).
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: H75106
R:Anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: H75106
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-186 <KAW>
A:Cross-references: GB:AJ248285; GB:AL096836; NID:95458067; PIDN:CAB49569.1; PID:9545
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0445
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 4.2%; Score 185; DB 2; Length 186;
Best Local Similarity 32.2%; Pred. No. 0.00015;
Matches 57; Conservative 33; Mismatches 61; Indels 26; Gaps 9;

OY 126 DLTTHAVDVANANEDLLHGGGLALALYKAGGFEL-----QESKQFVARYKVSAG 178
DB 11 DITFRKAEIVANANKYLFHGGGVAVAIKAAASGVSEYIRISKEMRKQIGR-DWIEHG 69
OY 179 EIAVT---GAGRUPCQIIHAGVPRMMEMDKQCTGKLORAIVSIINVIYKNTHTI 231
DB 70 EVAVTPEPLANNGKYYVHTVGPVW-DKR--KLELAIGLAKKA--DELGV 123
OY 232 KTVAIIPALSSGIFQFPLNCTKTIVETIIVSLQGRPMNSLKEIHLV--SNEDPTVA 286
DB 124 RSIAPFAISAGIYGCPLFEVVKTFKLVNFIKLS--AKNVADIVLVLYSERDYEVA 177

Search completed: May 13, 2003, 11:45:29
Job time : 54 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 11:30:37 ; Search time 27 Seconds

(Without alignments)
1311.881 Million cell updates/sec

Title: US-09-830-762-2

Perfect score: 4447
Sequence: 1 MDSMVGAAAYNEKSGRIT.....PMRPFAPWRCFASGSPVD 854

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	291.5	6.6	599	Y508_THEMA	Q9WYX8 thermotoga
2	291	6.5	182	YN83_SUITO	Q96XYS sulfobolus
3	288	6.5	177	Y599_SULSO	Q97WU4 sulfobolus
4	267.5	6.0	182	YB11_PYRAE	Q8ZXC3 pyrobaculum
5	235	5.3	176	YR59_LISMO	Q936S3 listeria mo
6	232	5.2	176	Y702_LISIN	Q93SX7 listeria in
7	228	5.1	183	Y189_ACISP	Q440Z0 alcaligenes
8	220	4.9	173	YGB2_ALCEU	Q3HXU7 pseudomonas
9	211	4.7	173	Y0J3_PSEAE	Q8ZGZ9 salmonella
10	203.5	4.6	179	YMDB_SALTY	Q59182 pyrococcus
11	203	4.6	190	YF13_PYRHO	Q8Y2K1 raietonia s
12	201.5	4.5	171	Y334_RALSO	Q9EY16 streptomyce
13	194	4.4	169	YSP0_STRCO	Q07733 mycobacteri
14	193	4.3	181	Y189_STRNO	P75918 escherichia
15	187.5	4.2	359	Y199_MCTU	Q9WYX3 pyrococcus
16	186.5	4.2	177	YMDB_ECOLI	Q9KHG2 streptomyce
17	185	4.2	186	Y445_PYRAB	Q88751 archaeoglob
18	184	4.1	177	Y189_STRGR	Q98542 rhizobium l
19	183	4.1	192	YF21_ARCFU	Q9HJ67 thermoplasma
20	182.5	4.1	176	Y72X_RHTLO	Q97AU0 thermoplasma
21	171	3.8	196	YB05_THEAC	Q93RG0 treponema m
22	168	3.8	186	Y189_THREO	Q67367 homo sapien
23	166	3.7	261	Y189_TREMD	Q67112 homiflex aeo
24	162.5	3.7	371	H2AY_HUMAN	Q02874 ratulus notv
25	159.5	3.6	165	Y987_AQUAE	Q05002 human.cocor
26	157	3.5	370	H2AY_RAT	Q9Y6E3 methanococ
27	156	3.5	4085	RRPA_CVH22	Q9Y6E3 methanococ
28	153	3.4	1169	SMC_METJA	Q9Y6E3 methanococ
29	151	3.4	194	YG48_AERPE	Q9Y6E3 methanococ
30	148	3.3	170	YMB8_DEIRA	Q9Y6E3 methanococ
31	147.5	3.3	2485	POLN_EEYV3	P36327 venezuelan
32	146	3.3	266	Y325_STAMM	Q99WQ1 staphylococ
33	139	3.1	2492	POLN_EEYV1	P27282 venezuelan

ALIGNMENTS

RESULT 1	ID	Y508_THEMA	STANDARD:	PRT:	599 AA.
AC	Q9WYX8:	Y508_THEMA			
DT	15-JUN-2002 (Rel. 41, Created)				
DT	15-JUN-2002 (Rel. 41, Last sequence update)				
DE	Hypothetical protein TM0508.				
GN	TM0508.				
OS	Thermotoga maritima.				
OC	Bacteria; Thermotogae; Thermotogae (class); Thermotogales;				
OX	NCBI_TaxID=2336;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MSR8 / DSM 3109;				
RX	MEDLINE=99287316; PubMed=10360571;				
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,				
RA	Hart D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,				
RA	McDonald L., Uterback T.R., Malek J.A., Linher K.D., Garrett M.M.,				
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,				
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,				
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser J.C.,				
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from				
RT	genome sequence of Thermotoga maritima."				
RL	Nature 399:323-329(1999).				
CC	-1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE UPF0189				
CC	FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
CC	EMBL: AE001727; A035593.1; -				
DR	TIGR: TM0508;				
DR	InterPro: IPR002589; A1PP.				
DR	InterPro: IPR003593; AAA_ATPase.				
DR	InterPro: IPR003959; AAA_ATPase-centr.				
DR	InterPro: IPR000862; RFGdomain.				
DR	Pfam: PF00004; AAA: 1.				
DR	Pfam: PF01661; A1PP: 1.				
DR	SMART: SM00506; A1PP: 1.				
DR	SMART: SM00382; AAA: 1.				
KW	Hypothetical protein; ATP-binding; Complete proteome.				
FT	NP_BIND 49 56				
FT	NP_BIND 49 56				
FT	DOMAIN 422 599				
FT	SEQUENCE 599 AA; 67417 MW; BOFBSCB87155BC7 CRC64;				

Query Match 6.6%; Score 291.5; DB 1; Length 599;
Best Local Similarity 31.9%; Pred. No. 9.3e-11;
Matches 80; Conservative 47; Mismatches 91; Indels 33; Gaps 7;

OY 57 GENYSWOIPIINHNDP-----KILKNN-----EROLCEVLONKEGCIJSTLWSPV 99
 DB 359 GEGYLY--PHDFGFGFVKTNYLPEKLENEVIFQPKRGVEEELERLKLW-----PE 408
 OY 100 OEGSKSLQVFRKMLTPRIELSVKWKDDLTTHAVDAVYVNAANEDLLHGGGLALALYKAGCF 159
 DB 409 KYGGSMAEYKKELEYGKKRTIRYKGDITREEDAIYVNAANEYKHGGGVALVRAGGS 468
 OY 160 EIOESKOPFARYKGSVAGELAVTAGRLPCKQIITHAVGPRMWMEDKQCTGKLORAIYS 219
 DB 469 VIOESDRIVQERGRVPTGEAVVTSAGKLAKKYVITHVGPW----RGSGDEDELLYKA 524
 OY 220 ILAVVITKKN--THITVAIPALSSGIFQFPPLNLCTKTIVETIRVSLQCKPMKSNKEIHLV 278
 DB 525 VYNALLLAHLELKLSISMPSALSTGIFGPKERAVGIFSKAIRFDIDPH--DTVLEIRIC 583
 OY 279 SNEDEPTVAAFK 289
 DB 584 NIDETTKIFE 594

RESULT 2
 YN83_SULTO STANDARD; PRT; 182 AA.
 AC 096XY5:
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ST2383.
 GN ST2383.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=111955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JCM 10545 / 7;
 RX MEDLINE=21456156; PubMed=11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 RA Sakine M., Baba S.-I., Anai A., Kosugi H., Hosoyama A., Fukui S.,
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 RA Yoshizawa T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A.,
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
 RA Oshima T., Kikuchi H.;
 RT "Complete genome sequence of an aerobic thermophilic
 RT Crenarchaeon, Sulfolobus tokodaii strain7.";
 RL DNA Res. 8:123-140(2001).
 CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AP000989; BAB67492.1; -
 DR InterPro: IPR002589; Alpp. 1.
 DR Pfam: PF01661; Alpp. 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 182 AA; 20183 MW; 1E0A072586AF8947 CRC64;
 Query Match 6.5%; Score 291; DB 1; Length 182;
 Best local Similarity 38.5%; Pred. No. 1.9e-11;
 Matches 62; Conservative 37; Mismatches 46; Indels 16; Gaps 4;

DB 63 VAVTSAGKLAKKAVIYHAVGPRY-----GIEGEKELEAIRNALRKA--BELKLSSIALP 114
 OY 238 ALSSGIFQFPPLNLCTKTIVETIRVSLQCKPMKSNKEIHLV 278
 DB 115 AISTGITGYPEICAEKMKVY-----KEEYTFEKLHNTI 149

RESULT 3
 YS99_SULSO STANDARD; PRT; 177 AA.
 AC 097JUN4:
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein SSO2899.
 GN SSO2899.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aweyer M.J., Chan-Welner C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moers A., Erauso G., Fletcher C., Gordon P.M.K.,
 RA Helkamp-de Jong I., Jaffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doollittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.,
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE006882; AA43008.1; -
 DR InterPro: IPR002589; Alpp.
 DR Pfam: PF01661; Alpp. 1.
 DR SMART: SM00506; Alpp. 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 177 AA; 19445 MW; 8353685B14FECA96 CRC64;
 Query Match 6.5%; Score 288; DB 1; Length 177;
 Best local Similarity 40.7%; Pred. No. 2.7e-11;
 Matches 59; Conservative 33; Mismatches 43; Indels 10; Gaps 3;

OY 118 IELSVKMDLTTHAVDAVYVNAANEDLLHGGGLALALYKAGFELQESKQFVARYKVS 177
 DB 8 LEYLINGDITTELEADAIVNANASYLQHGCGVAIVAKKGYIIQKSDSEYKFGFVPV 67
 OY 178 GEIAVTAGRLPCKQIITHAVGPRMWMEDKQCTG--KLORAIYSILVYVYKNTHTKIVA 235
 DB 68 GEVAVTSAGKLAKKAVIYHAVGPRY-----GIEGDEKLESAIFRSL--LAADELSSSIA 119
 OY 236 IPALSSGIFQFPPLNLCTKTIVETIR 260
 DB 120 MPALSTGTYGYPEICARIMANVYLK 144

RESULT 4
 YB11_PYRAE STANDARD; PRT; 182 AA.
 ID YB11_PYRAE
 AC 082XTJ3;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)

[illegible]

CC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PAOI;
 RX MEDLINE=2043733; PubMed=10964043;
 RA Slover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzer S.N., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.S., Wu Z., Paulsen I.T.,
 RA Kelter J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE004789; AGO7081.1; -
 DR InterPro: IPR002589; A1PP.
 DR Pfam: PF01661; A1PP; 1.
 DR SMART: SM00506; A1PP; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 173 AA; 18219 MW; D94A49C6FD208BF8 CRC64;
 Query Match 4.7%; Score 211; DB 1; Length 173;
 Best Local Similarity 35.2%; Pred. No. 1.6e-06;
 Matches 57; Conservative 28; Mismatches 65; Indels 12; Gaps 5;
 OY 119 ELVWKDILTHAVDAVYVNAANEDLLHGGGLALVYKAGFEIOESKQFVARYGVSAG 178
 DB 3 EVRWGGITRLAVDAIVANANSSLLGGGVDGAIHRAAGAEIVACRL--HG-CKTG 58
 OY 179 ELVATGAGRLPCKOIITHAVGPRMWMWDKQGTGKLO-RAIVSLINVYKTHIKTVAIIP 237
 DB 59 EAKITRGRLPAHVIHVGVPW---RGDNGEAEELASCYRSLALAEONAGASVAP 114
 OY 238 ALSSGIFQPLNCTKTIYETIRVSLQCKPMMSNKEIHLVS 279
 DB 115 AISCIGYPLEQAAIAVEEV---CRQRPASSLEIIVVA 153
 RESULT 10
 YMDB_SALTY
 ID YMDB_SALTY STANDARD; PRT; 179 AA.
 AC 08Z029; 08Z7M1;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Hypothetical protein ymdb.
 GN YMDB OR STM1147 OR STY1184.
 OS Salmonella typhimurium, and
 OS Salmonella typhi.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 OX NCBI_TaxID=602, 601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium

RT LT2.";
 RL Nature 413:852-856(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhi; STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.F.G., Sebaiha M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Fellwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Goara P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Bartell B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF008749; AAL20077.1; -
 DR EMBL: AL627269; CAD08271.1; ALT_INT.
 DR StyGene; SG77777; ymdb.
 DR InterPro: IPR002589; A1PP.
 DR Pfam: PF01661; A1PP; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 179 AA; 19196 MW; 05AE468EF0CD9315 CRC64;
 Query Match 4.6%; Score 203.5; DB 1; Length 179;
 Best Local Similarity 32.9%; Pred. No. 4.8e-06;
 Matches 46; Conservative 31; Mismatches 60; Indels 3; Gaps 2;
 OY 120 LSVWKDILTHAVDAVYVNAANEDLLHGGGLALVYKAGFEIOESKQFVARYGVSAGE 179
 DB 5 LQVIGDITQISVDAIVANANSSLLGGGVDGAIHRAAGAPLDDCKLIRQOGGECQIGH 64
 OY 180 IAVTCAGRLPCKOIITHAVGPRMWMWDKQGTGKLORAIVSLINVYKTHIKTVAIIPAL 239
 DB 65 AVTPACKLSAKAVIHVGPWRGGEHQEAE-LLEAVRNCL--LLAANFRSIAFPAL 121
 OY 240 SSGIFQPLNCTKTIYETI 259
 DB 122 STGVYGYPRQAQAAEVAVRTV 141
 RESULT 11
 YF13_PYRHO
 ID YF13_PYRHO STANDARD; PRT; 190 AA.
 AC 059182;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Hypothetical protein PH1513.
 GN PH1513.
 OS Pyrococcus horikoshii.
 CC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 CC Pyrococcus.
 OX NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OT3;
 RX MEDLINE=98344137; PubMed=9679194;
 RA Kanarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
 RA Yamamoto S., Sekine W., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
 RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,

RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
 RA Masuuchi Y., Shizuya H., Kikuchi H.,
 RT "Complete sequence and gene organization of the genome of a hyper-
 RT thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.",
 RL DNA Res. 5:55-76(1998).
 CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC -----
 DR EMBL: AP000006; BAA30621.1; -
 DR InterPro: IPR002589; Alpp.
 DR Pfam: PF01661; Alpp; 1.
 DR SMART: SM00506; Alpp; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 190 AA: 21106 MW: 85DE0761BE3A684 CRC64;
 Query Match 4.6%; Score 203; DB 1; Length 190;
 Best Local Similarity 34.2%; Pred. No. 5,6e-06;
 Matches 66; Conservative 38; Mismatches 59; Indels 30; Gaps 12;
 QY 124 KDLTTHAVDAVYVNAANEDLLHGGGLALALVAKGFEIOE---SKQFVAR-YGK--VSA 177
 DB 9 RCDITXRRAAIYVNAANKRYLEHGGGVAYAKAAGVSEYTRISKEMRROLCKDWIEH 68
 QY 178 GELAVNGAGRLP---CKQIHAVGPR---WMEMDKGCGTCKLORAVSLNVIYKNTH 230
 DB 69 GEVAVVPPMKLKGKNGVYVHTVGPYCGVSK-DKE---EKRLALALGALKKA--DELG 122
 QY 231 IKTVAPALSSGIFQFPLNCTKTIVETIRVSLQCKPMNSLNKEIHLV--SNEDPTVAAF 288
 DB 123 VKSIAPFALISAGIYCGPLKEVTRFEKVVAEFLK--VANHVKEVYLVLXSEBD-----Y 174
 QY 289 KAASEFI-LCKSE 300
 DB 175 KLALETIGLCDND 187
 RESULT 12
 Y334_RALSO STANDARD; PRT; 171 AA.
 AC 08Y2K1;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein RSC0334.
 GN RSC0334 OR RSC03301.
 OS *Ralstonia solanacearum* (Pseudomonas solanacearum).
 CC *Ralstonia solanacearum*; beta subdivision; *Ralstonia* group;
 CC Bacteria; Proteobacteria; beta subdivision;
 CC *Ralstonia*.
 RT NCBI_TaxId=305;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Arlat M., Billault A., Brotier P., Camus J.C., Catrolicio L.,
 RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Laveit M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
 RA Weisenbach J., Boucher C.A.,
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*,"
 RL Nature 415:497-502(2002).
 CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC -----
 DR EMBL: AL646058; CAD13862.1; -
 DR InterPro: IPR002589; Alpp.
 DR Pfam: PF01661; Alpp; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 171 AA: 18018 MW: 7DC99AF217882657 CRC64;
 Query Match 4.5%; Score 201.5; DB 1; Length 171;
 Best Local Similarity 34.2%; Pred. No. 6e-06;
 Matches 52; Conservative 24; Mismatches 61; Indels 15; Gaps 4;
 QY 113 MLTPRIELSVWKDGLTTHAVDAVYVNAANEDLLHGGGLALALVAKGFEIOESKQFVAR 172
 DB 1 MPPIPTVLRALRADITTLACDAIVNANSLDGGGVDAIHRNAGEELLEACR---ALH 57
 QY 173 GKVSAGEIATVAGRLPCQKQIHAVGPRWEMDKGCTGKLRVSIILNY---VIYN 228
 DB 58 G-CRTGQAKITPGLLPARYIHTVGPVWR-----GGRDEAALAACVRSNLALAKO 109
 QY 229 THIKTVAPALSSGIFQFPLNCTKTIVETIR 260
 DB 110 HDVKTIAFPCLSTGVYGFPPQLAPIAVATVR 141
 RESULT 13
 Y330_STRCO STANDARD; PRT; 169 AA.
 AC 09ZBG3;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein SC06450.
 GN SC06450 OR SC985.17.
 OS *Streptomyces coelicolor*.
 CC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 CC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
 CC NCBI_TaxId=1902;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21966410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Horsby T., Howarth S.,
 RA Huang C.-H., Krieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.,
 RT "Complete genome sequence of the model actinomycete *Streptomyces*
 RT *coelicolor* A3(2).",
 RL Nature 417:141-147(2002).
 CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC -----
 DR EMBL: AL035206; CAA22759.1; -
 DR InterPro: IPR002589; Alpp.
 DR Pfam: PF01661; Alpp; 1.
 DR SMART: SM00506; Alpp; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 169 AA: 17767 MW: 86965DB3336E4740 CRC64;

Query Match 4.4%; Score 194; DB 1; Length 169;
 Best Local Similarity 30.5%; Pred. No. 1.7e-05;
 Matches 53; Conservative 35; Mismatches 72; Indels 14; Gaps 5;

OY 120 LSWMKDDLTTHAVDAVYVNAANEDLLHGGLALALVKAGFEIOESKOVARY-OK-VSA 177
 DB 4 ITLVGGDITRQADALVYVNAANSSLLGGGGVGAHRRGGPALLAECCRLRAGHLGGLPT 63
 OY 178 GEIAVTCAGRLPCKOIITHAVGPRMWM-DKOCCTGLORAIYSILNYVYKTHIKTVAI 236
 DB 64 GRAVATTAAGDLDARVYIHVGPVMSATEDRSGLLASCYRESLRTAD-----ELGARTVAF 118
 OY 237 PALSSGIFQPLNLTCTIVETIRYSLOQKPMNSLKEIHLVSNEDPVAAFKA 290
 DB 119 PAISTGVYRMPMDARIAIVETVATV-----KTSVTEIRFVLFDARAYAEFAA 166

RESULT 14
 Y189_STRNO STANDARD; PRT; 181 AA.
 ID Y189_STRNO
 AC Q9EXY6;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical 19.0 kDa protein in sno 5' region (ORF7).
 OS Streptomyces nogalater.
 OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;
 OC Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=38314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 27451;
 RX MEDLINE=21539456; PubMed=11683270;
 RA Torkell S., Kumari T., Palmu K., Meentsaelae P., Hakala J.,
 RA Ylilomko K.;
 RT "The entire nogalater biosynthetic gene cluster of Streptomyces
 RT nogalater: characterization of a 20-kb DNA region and generation of
 RT hybrid structures.";
 RL Mol. Genet. Genomics 266:276-288 (2001).
 CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AF323753; AAG42849.1; -;
 CC InterPro: IPR002589; Alpp.
 CC Pfam: PF01661; Alpp; 1.
 CC SMART: SM00506; Alpp; 1.
 CC Hypothetical protein.
 CC KW Hypothetical protein.
 CC SEQUENCE 181 AA; 19048 MW; C9184D0A793F4CA CRC64;

Query Match 4.3%; Score 193; DB 1; Length 181;
 Best Local Similarity 32.9%; Pred. No. 2.2e-05;
 Matches 48; Conservative 28; Mismatches 62; Indels 8; Gaps 4;

OY 120 LSWMKDDLTTHAVDAVYVNAANEDLLHGGLALALVKAGFEIOESKOF-VARYK-VSA 177
 DB 4 ITLVGGDITRQADALVYVNAANSSLLGGGGVGAHRRGGPALLAECCRLRAGHLGGLPT 63
 OY 178 GEIAVTCAGRLPCKOIITHAVGPRMWM-DKOCCTGLORAIYSILNYVYKTHIKTVAI 236
 DB 64 GRAVATTAAGDLDARVYIHVGPVMSATEDRSGLLASCYRESLRTAD-----ELGARTVAF 118
 OY 237 PALSSGIFQPLNLTCTIVETIRVS 262
 DB 119 PALSTGVYRMPMDARIAIVETVRTT 144

RESULT 15
 Y199_MYCTU STANDARD; PRT; 359 AA.
 ID Y199_MYCTU
 AC 007733; ORYJ07;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein RV1899C.
 GN RV1899C OR MT1950 OR MYCY180.19.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=968295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sutcliffe J.E., Taylor K., Whitehead S., Barrall B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M., Hatt D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A.L., Utterback T., Weidman J., Khouri H., Gill J., Minkula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE UPF0189
 CC FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: 297193; CAB10035.1; ALT_INIT.
 CC EMBL: AE007050; AAK46221.1; -;
 CC TIGR: MT1950; -;
 CC DR TubercuList: RV1899C; -;
 CC InterPro: IPR002589; Alpp.
 CC Pfam: PF01661; Alpp; 1.
 CC SMART: SM00506; Alpp; 1.
 CC Hypothetical protein; Complete proteome.
 CC KW DOMAIN 184 UNKNOWN.
 CC FT DOMAIN 185 359 UPF0189.
 CC FT CONFLICT 76 76 A -> S (IN REF. 2).
 CC FT CONFLICT 126 126 C -> S (IN REF. 2).
 CC SEQUENCE 359 AA; 36936 MW; F73BBB4EF5B1426 CRC64;

Query Match 4.2%; Score 187.5; DB 1; Length 359;
 Best Local Similarity 29.0%; Pred. No. 0.00013;
 Matches 56; Conservative 32; Mismatches 74; Indels 31; Gaps 5;

OY 118 IELSVKDDLTTHAVDAVYVNAANEDLLHGGLALALVKAGFEIOESKOFVARYGVSA 177
 DB 190 IELVHQAQVTKLEDAITINANTRLRHNGVYAAATARAGGPELOESIE-----KAPIGL 245

OY 178 GEIAVGTAGRLPCKOIHAV-----GPRMMEWDKOGCTGKLORAIIVSILNVYIKNTHIK 232
Db 246 GEAVETTAGDMPARYVTHAATMELGPTSGEITATAATLTKA-----DELGCR 295
OY 233 TVALPALSSGIFPFPLNCTKTIVETIRVSLQCKPMSNLKEIHLVSNEDPTVAAPKAAS 292
Db 296 SLALVAFGTGVGFPDPAARLWGAVRRHRPG-----SLQRVVFAVHGDAAERAFSNA- 349
OY 293 EFLGKSELGOET 305
Db 350 -----IOAGEDT 356

Search completed: May 13, 2003, 11:42:45
Job time : 31 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 11:38:38 ; Search time 100 Seconds
(without alignments)
1759.642 Million cell updates/sec

Title: US-09-830-762-2
Perfect score: 4447
Sequence: 1 MDESMVAGAAAYNEKSGRIT.....PMRPFADHPWRGFGASGSPVD 854

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4438	99.8	854	09BZL9	Q9BZL9 homo sapien
2	4220.5	94.9	819	09BZL8	Q9BZL8 homo sapien
3	2453	55.2	830	11_099LFP9	Q99LFP9 mus musculu
4	1747	39.3	330	08TCF3	Q8TCF3 homo sapien
5	610	13.7	1023	4_09ULF2	Q9ULF2 homo sapien
6	420	9.4	556	4_09NV60	Q9NV60 homo sapien
7	256	5.8	419	4_09H9X9	Q9H9X9 homo sapien
8	234.5	5.3	175	16_08RB30	Q8RB30 thermoanaer
9	231	5.2	298	13_08QGB0	Q8QGB0 oncorhynchu
10	230.5	5.2	424	11_08VDT6	Q8VDT6 mus musculu
11	215	4.8	183	17_08UOP9	Q8UOP9 pyrococcus
12	211.5	4.8	1025	4_096K72	Q96K72 homo sapien
13	207	4.7	672	4_08WV05	Q8WV05 homo sapien
14	205	4.6	393	4_096CH7	Q96CH7 homo sapien
15	203	4.6	363	4_09HR89	Q9HR89 homo sapien
16	200.5	4.5	654	11_08RI33	Q8RI33 mus musculu

17	197	4.4	606	11_08RI09	Q8RI09 mus musculu
18	195.5	4.4	227	4_08IAJ3	Q8IAJ3 homo sapien
19	190	4.3	243	4_09B069	Q9B069 homo sapien
20	190	4.3	325	4_09U986	Q9U986 homo sapien
21	189	4.3	359	4_09IAP7	Q9IAP7 homo sapien
22	186.5	4.2	175	16_08RHQ2	Q8RHQ2 fusobacteri
23	182	4.1	195	17_08RQD0	Q8RQD0 methanosarc
24	182	4.1	243	11_0922B1	Q922B1 mus musculu
25	178	4.0	193	10_0228P5	Q228P5 arabidopsis
26	177	4.0	193	10_0949P6	Q949P6 arabidopsis
27	176	4.0	499	12_08QUT8	Q8QUT8 infectious
28	171.5	3.9	182	17_08WT2	Q8WT2 methanopyru
29	168	3.8	369	13_0933Z6	Q933Z6 gallus gall
30	166	3.7	369	11_091VZ2	Q91VZ2 mus musculu
31	165.5	3.7	4383	12_08VW7	Q8VW7 mus musculu
32	161.5	3.6	372	4_09HBP3	Q9HBP3 bovine coro
33	161.5	3.6	372	4_09HBP1	Q9HBP1 homo sapien
34	160.5	3.6	2447	12_09WUD0	Q9WUD0 venezuelan
35	160	3.6	203	5_017432	Q17432 caenorhabdi
36	159.5	3.6	4383	12_091A28	Q91A28 bovine coro
37	159.5	3.6	7094	12_091A29	Q91A29 bovine coro
38	159.5	3.6	7094	12_08VA39	Q8VA39 bovine coro
39	159.5	3.6	7094	12_08VA39	Q8VA39 bovine coro
40	157.5	3.5	1181	5_09XZ37	Q9XZ37 drosophila
41	157.5	3.5	1181	5_09VBP3	Q9VBP3 drosophila
42	156.5	3.5	372	11_09QZQ8	Q9QZQ8 mus musculu
43	156	3.5	258	4_09H610	Q9H610 homo sapien
44	156	3.5	371	11_009140	Q09140 rattus norv
45	156	3.5	598	4_09NT13	Q9NT13 homo sapien

ALIGNMENTS

RESULT 1

ID	Q9BZL9	PRELIMINARY;	PRT;	854 AA.
AC	Q9BZL9;			
DT	01-JUN-2001 (TREMBL)	17, Created)		
DR	01-JUN-2001 (TREMBL)	17, Last sequence update)		
DT	01-DEC-2001 (TREMBL)	19, Last annotation update)		
DE	B aggressive lymphoma long isoform.			
GN	BAL.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20563954; PubMed=11110709;			
RA	Aguiar R.C.T., Yakushiji Y., Kharbanda S., Salgia R., Fletcher J.A.,			
RA	ShiPP M.A.;			
RT	"BAL is a novel risk-related gene in diffuse large B-cell lymphomas			
RL	Blood 96:4328-4334(2000).			
DR	EMBL; AF307338; AAK07558.1;			
DR	InterPro; IPR002589; A1pp.			
DR	Pfam; PF01661; A1pp; 2.			
DR	SMART; SM00506; A1pp; 2.			
SO	SEQUENCE 854 AA; 96283 MW; E80663868ECDECD3 CRC64;			
Query Match	99.8%; Score 4438; DB 4; Length 854;			
Best Local Similarity	99.9%; Pred. No. 4.5e-301;			
Matches	853; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			

Oy	1	MDESMVAGAAAYNEKSGITSLIPKQVFAQIPQWFKGNTPECLPYKSGETGALGNY 60
Db	1	MDESMVAGAAAYNEKSGITSLIPKQVFAQIPQWFKGNTPECLPYKSGETGALGNY 60
Oy	61	SMQPIINNDPRILKNNRQICEVLONKFGCSTLVSVOGNSKSLQVFRKMLTPREL 120
Db	61	SMQPIINNDPRILKNNRQICEVLONKFGCSTLVSVOGNSKSLQVFRKMLTPREL 120

Db 144 KNKHTNNEKLLFHTSSDSITQINNHGFSRYAGTHGAICNGSYFAVNSSYSARGVSKA 203

QY 753 SA-ADKLIYFEAEVLGFCQGHPLNIVPPPLSPGA-IDGHSDVDNVSSPEFVIFSG 810
 DB 204 DAQGNKRMVYL-ARVLVGDYTGQAGLIVPPAKPSGKADLDYSDVNTSNPTFVIFSD 261
 QY 811 MQALPOTL 818
 DB 262 VQAYPEFL 269

RESULT 10

OSVDT6 PRELIMINARY: PRT: 424 AA.
 ID 08VDT6
 AC 08VDT6
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical 48.4 kDa protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.
 RL Submitted (TAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC021340; AAI21340.1;
 DR InterPro; IPR004170; WME_dom.
 DR Pfam; PF02825; WME; 1.
 KM Hypothetical protein.
 SQ SEQUENCE 424 AA; 48423 MW; D71DFB740F315C07 CRC64;

Query Match 5.28; Score 230.5; DB 11; Length 424;
 Best Local Similarity 22.0%; Pred. No. 7.7e-08;

Matches 92; Conservative 70; Mismatches 168; Indels 89; Gaps 13;

QY 484 AKRSKMLSNVY-SVPOSTREKRENGLEAR--SPALNMGFVNEEYEAHWIQTLSL 540
 DB 16 AKOSVMSKIASFLGFPKQASPKNTLVLEKKIEHTFOVCGSVDSYKNTLSMLKELTK 75
 QY 541 -QNHIIENNHILYLGKREHDILOLKTSSVTEIISPGRELEEGARADLIEVVM 599
 DB 76 EQLSYTNDECVDFDEMEYEKLEIQRK--ELNITIEENQKTSIQVSGISRDVIKARDE 133
 QY 600 IEDMLCKVQ-----EEMARKKEGLMRSGLQMTIQ 630
 DB 134 IEGIKSIRLAKEKESQADYISTYEMQYIDKNITOCFDKMTNKLLEVAMKAKKDTVVQ 193
 QY 631 OKTOD--EMKEN-----IIFLKCVPPTQELLDQKQF----- 661
 DB 194 IHNDFVVDLSTNTATAPOGOTFYVQRVYKAEALIPANWSMKODKLLVNLQTSDEYN 253
 QY 662 -----EKCG-IQVLYVEKIDNEVLAAPORKKRMEKILHROPVSHRLFOQVYPOFCN 713
 DB 254 MVASAFQTCSDFEIKERIQNPALMRRYQAYKKGMDKNGVNRNKKHLFHTGEASSLP 313
 QY 714 VVCVGVQRMSTCDPKRYGAGITFTKNKLAKAKKISADKLI-----YVEAE 765
 DB 314 QLSNGSFRSAGKNATAYGKGTPE-----AVKA-SYSACDYYSRDPTNGRKYMYVR 365
 QY 766 VLTGFCGAGHLNIVPPPLSP-GAIDGHSDVDNVSSPEFVIFSGMQALPQVLTMTQ 823
 DB 366 VLTGNTYNGNASLIVPPSRDPQNAADLYDTYTDNDKNSITVIVYDQTYFEYLITRRQ 424

RESULT 11

OSVDT6 PRELIMINARY: PRT: 183 AA.
 ID 08VDT6
 AC 08VDT6
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein PF1536.
 GN PF1536.

OS Pyrococcus furiosus.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=2261;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VCI / DSM 3638 / ATCC 43587 / JCM 8422;
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
 RT "The complete sequence of the Pyrococcus furiosus genome";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE010254; AAL81660.1;
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 183 AA; 19937 MW; 9F3C2F7EAF525063 CRC64;

Query Match 4.8%; Score 215; DB 17; Length 183;

Best Local Similarity 36.5%; Pred. No. 2.5e-07;

Matches 65; Conservative 32; Mismatches 57; Indels 24; Gaps 10;

QY 120 LSVWKDDLTTHAVDAVYVNAANEDLLHGGGLALALVKGFEIQEE--SKQV-ARYK- 174
 DB 2 IKVYKGDITKFRADAIYVANKYLEHGGGVAIAKAAAGDVREYIRISKEAREQLGKD 61
 QY 175 -VSAGEIYVAGARLP---CKQIHAVP---RWMEKDGCGCTGKIORIVSILNVIY 226
 DB 62 WIDHGEVYVPTQLQEKNGKYVHTVGYCGSWDE-DKK---SKUKLILGALKKA-- 115
 QY 227 KNTIKTVAIPALSSGIFOPNLCTKTIYETIRVSLGKPKMSNLEIHLV--SNED 282
 DB 116 DELGVKSIAFPALPASIGIYCGPLEKVEYTFEVVKEFL---PSAKSLREYFLVLYSQED 170

RESULT 12

OSVDT6 PRELIMINARY: PRT: 1025 AA.
 ID 096K72
 AC 096K72
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE CDNA FLJ14464 f15, clone MAMMA1000309.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MAMMARY GLAND;
 RA Isegaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nagatsuna M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugakura M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
 RA Ninomiya K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK027370; BAB55067.1;
 SQ SEQUENCE 1025 AA; 109884 MW; 87C4DBB8F42DB84C CRC64;

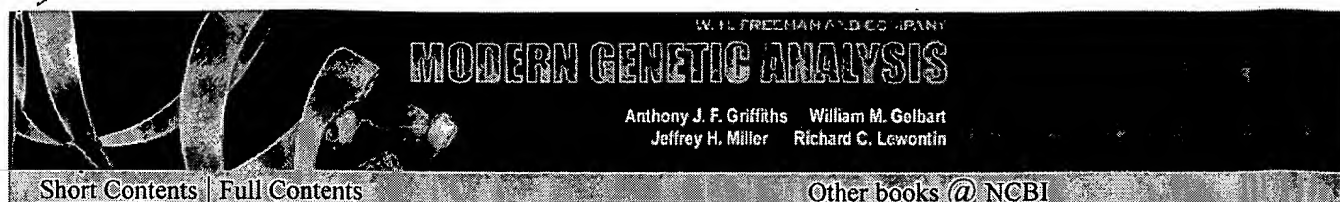
Query Match 4.8%; Score 211.5; DB 4; Length 1025;

Best Local Similarity 21.2%; Pred. No. 6.5e-06;

Matches 194; Conservative 115; Mismatches 358; Indels 247; Gaps 41;

QY 77 NEROL---CEVIONKFGCISTLVSPVQ-----BGSKSLQVFRKMLTPR- 117
 DB 195 NERRSGGPLEDLQRLPGLGTVAASFQOVAVERVLQOEHRLOSSELSLVPHYDLEPPE 254
 QY 118 -IELSVKQDILTTHAVDAVYVNAANEDLLHGGGLALALVKGGEPI--QESKQFVA--RY 172
 DB 255 LAENTSGDHPSTGQPRATYHA-----LLRTGGVLTALQAGVYVWGSGEEOGSGASLRT 310
 QY 173 GKSAGE-IATVGAGRLPCKQ-IIHAVGPRMMEWDKCGCTGKIORAIVSILNVIYKNT 230
 DB 311 GPVVGRCIMTGTSGGEPGSGTSLRTGP-----MGSILQAMQ-----VSSNMGSLEH 359

[illegible]



Navigation

About this book

11.

Applications of Recombinant DNA Technology

Introduction

In Vitro

Mutagenesis

→ RFLP Mapping

Reverse Genetics

Expressing Eukaryotic Genes in Bacteria

Recombinant DNA Technology in Eukaryotes

Gene Therapy

Using Recombinant DNA to Detect Disease Alleles Directly

Summary

Concept Map

Solved Problem

Solution

Solved Problem

Problems

Figures

Figure 11-2.

Mapping a gene...

Figure 11-3.

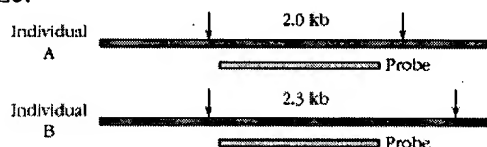
The detection and...

Search

Modern Genetic Analysis → 11. Applications of Recombinant DNA Technology

RFLP Mapping

In Chapter 10, we learned that, if a cloned DNA fragment is used as a probe of genomic DNA that has been cut with a restriction enzyme, then the probe will bind to one or more genomic fragments. For example, if the restriction enzyme used does not cut within the chromosomal region encompassed by the cloned fragment, then the probe should bind to one fragment flanked by restriction sites on each side. Since the DNA of chromosomes within a species is generally homologous, it might be expected that a constant-sized genomic fragment will be bound in all individuals. However, when probes are used in this way, the bound fragments are often found to be of different sizes in different individuals. The explanation is that a given restriction site is not always found in all individuals. The absence of a site is usually caused by a single nucleotide difference that is most likely biologically neutral. Hence, for example, if a probe binds a 2-kb fragment in individual A of a haploid species and it binds a fragment of 2.3 kb in individual B, the reason is usually that one of the sites that flanked the 2-kb fragment is missing in B, and the next site is 0.3 kb away, making the hybridized fragment 2.3 kb in size.



The presence and absence of the restriction site can be treated as two alleles that can be thought of as + and - alleles. The presence of the + in some individuals in the population and the absence (-) in others generates a **restriction fragment length polymorphism, or RFLP**. (In the case just discussed, there was a dimorphism—two “morphs,” one short and one long.) Geneticists were surprised to discover that RFLPs are quite common in populations and that a large proportion of probes will detect one. RFLPs are identified by a rather hit-or-miss method of hybridizing panels of randomly cloned genomic fragments to genomic restriction digests of several different individuals in a family or a population. Because RFLPs are a relatively common type of variation in nature, this method succeeds in finding RFLPs in most cases.

The significance of RFLPs is threefold. First, if an individual is heterozygous for two morphs of an RFLP, this heterozygous “locus” can be used as a marker in chromosomal mapping. Although at first the locus of the RFLP is not necessarily known, as more and more RFLPs are found, they can be mapped in relation to gene loci and in relation to other RFLP loci, and their positions gradually saturate the genetic map. The RFLPs are not biologically significant in most cases, but they can be used to map interesting genes and act as positions from which these

THIS PAGE BLANK (USPTO)

Go

- ☒ This book
☐ All books
☐ PubMed

genes can be cloned by positional cloning.

Second, in an extension of mapping analysis, RFLP alleles (morphs) can be used as diagnostic tools. For example, in a family with a record of a certain disease, if it can be established that the people who have the disease also carry a specific allele of an RFLP, then this fact suggests not only that the RFLP locus is linked to the disease gene locus, but furthermore that the specific RFLP allele is in *cis* arrangement with the disease allele. Hence the RFLP allele becomes a diagnostic marker for the disease, and this information can be used in genetic counseling.

Third, RFLPs can be used to measure genetic divergence between different populations or related species. The restriction-site difference is effectively a DNA difference, so a measure of the total number of RFLP differences represents a measure of genetic difference. Hence RFLPs are important in studies of evolution.

RFLP mapping is often performed on a defined set of strains or individuals that become “standards” for mapping that species. For example, in the fungus *Neurospora*, two wild-type strains, Oak Ridge and Mauriceville, are known to show many RFLP differences, so these strains have become standards used in RFLP mapping. The RFLPs can be mapped relative to one another or to genes of known phenotypic expression. For example, let *ad* stand for an allele for adenine requirement, and 1 and 2 stand for RFLP loci with either the Oak Ridge (OR) or Mauriceville (M) “alleles.” A cross can be made of the type $ad \cdot 1^{OR} \cdot 2^{OR} \times ad^{+} \cdot 1^{M} \cdot 2^{M}$. Progeny are tested for all three loci. Adenine requirement is tested by inoculating strains on medium lacking adenine, and the RFLP alleles are tested by probing with the relevant probes. Recombinant frequencies are calculated in the usual way. Most mutants in *Neurospora* have been induced in Oak Ridge wild-type strains, so it is a simple matter to map the mutant alleles to RFLPs simply by crossing the mutant Oak Ridge strain to the wild-type Mauriceville strain. An example of mapping a phenotypic mutant by using RFLP markers is shown in [Figure 11-2](#).

Similar standard strains have been established in other organisms. An analogous approach has been used in human genome mapping by collecting DNA from a defined set of individuals in 61 families with an average of eight children per family and making this DNA available throughout the world to provide a standard for RFLP mapping.

[Figure 11-3](#) shows an example of linkage of a human disease allele to an RFLP locus and the potential for using this information in diagnostics. Because of the close linkage, future generations of persons showing the RFLP morph 1 can be predicted to have a high chance of inheriting the disease allele *D*. This sort of predictive power can be used in prenatal diagnoses of the genotypes of fetuses, with the use of amniocentesis or chorionic villus sampling (considered later in this chapter).

THIS PAGE BLANK (USPTO)

MESSAGE

RFLPs provide useful molecular marker loci for chromosome mapping and for diagnosis of human disease alleles.

It is worth comparing the process of making a restriction map (*restriction mapping*, pages 327–329) with the process of *RFLP mapping*. Restriction maps are based on *physical* analysis of DNA, whereas RFLP maps are based on *recombination* analysis of matings. Note also that restriction mapping is based on restriction sites with no variation, whereas RFLP mapping is based on restriction-site variation between homologous chromosomes. Most restriction maps are short-range (fine-scale) maps, although long-range maps can be constructed with rare-cutting restriction enzymes. In contrast, RFLP mapping generally produces long-range (coarse-scale) maps. RFLP mapping of whole genomes will be covered in detail in Chapter 12. ♦ [TOP](#)

© 1999 by W. H. Freeman and Company. All rights reserved.

THIS PAGE BLANK (USPTO)